

Preservação de Longo Prazo de
Conhecimento Codificado
Proposição de um Framework

Eloi Juniti Yamaoka

Eloi Juniti Yamaoka

**PRESERVAÇÃO DE LONGO PRAZO DE
CONHECIMENTO CODIFICADO:
PROPOSIÇÃO DE UM *FRAMEWORK***

Tese submetida ao Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina para obtenção do Grau de Doutor em Engenharia e Gestão do Conhecimento.

Orientador: Fernando Alvaro Ostuni Gauthier

Co-orientador: Gregório Jean Varvakis Rados

Florianópolis
2014

Ficha de identificação da obra elaborada pelo autor
através do Programa de Geração Automática da Biblioteca Universitária
da UFSC.

Yamaoka, Eloi Juniti

Preservação de Longo Prazo de Conhecimento Codificado:
Proposição de um Framework / Eloi Juniti Yamaoka;
orientador, Fernando Alvaro Ostuni Gauthier;
coorientador, Gregório Varvakis. – Florianópolis, SC, 2014.
202 p.

Tese (doutorado) – Universidade Federal de Santa
Catarina, Centro Tecnológico. Programa de Pós-Graduação em
Engenharia e Gestão do Conhecimento.

Inclui referencias

1. Engenharia e Gestão do Conhecimento. 2. Preservação Digital.
3. Objetos Digitais. 4. Avaliação Documental. I. Gauthier,
Fernando Alvaro Ostuni. II. Varvakis, Gregório. III Universidade
Federal de Santa Catarina. Programa de Pós-Graduação em
Engenharia e Gestão do Conhecimento. IV. Título.

Eloi Juniti Yamaoka

PRESERVAÇÃO DE LONGO PRAZO DE CONHECIMENTO
CODIFICADO: PROPOSIÇÃO DE UM *FRAMEWORK*

Esta Tese foi julgada adequada para obtenção do Título de “Doutor em Engenharia e Gestão do Conhecimento”, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina.

Florianópolis, 18 de fevereiro de 2014.

Prof. Gregório Jean Varvakis Rados, Dr.
Coordenador do Curso

Banca Examinadora:

Prof. Fernando Álvaro Ostuni Gauthier, Dr.
Orientador
UFSC

Prof^a. Helena de Fátima Nunes
Silva, Dr^a
UFPR

Prof. Renato Tarciso Barbosa
de Sousa Dr.
UnB

Prof. Márcio Matias Dr.
UFSC

Prof^a. Gertrudes Aparecida
Dandolini, Dr^a.
UFSC

Prof. Rogerio Cid Bastos, Dr.
UFSC

À Ilma, Letícia e Luísa

AGRADECIMENTOS

Meu agradecimento especial ao Professor Gauthier, que pacientemente orientou, aconselhou e transmitiu o conhecimento necessário.

Ao Professor Gregório, co-orientador e coordenador do curso.

Aos professores membros da banca de qualificação e defesa: Renato Tarciso Barbosa de Sousa, Helena de Fátima Nunes da Silva, Márcio Matias, Gertrudes Aparecida Dandolini, Rogério Cid Bastos e José Leomar Todesco.

Ao SERPRO, pelo apoio através do Programa de Educação Pós-Graduada.

Aos amigos da CETEC/SERPRO, especialmente ao Tiboni e José Maria pelo apoio e incentivo.

Aos professores do EGC, pelos ensinamentos.

Aos colegas do EGC, que embora tenha convivido muito menos do que gostaria, ficou a certeza que foi o início de muitas amizades perenes.

Ao Luiz, Severino e Cezar amigos de estudos e das jornadas para Florianópolis.

A Isamir, amiga nos estudos e no compartilhamento das angústias e alegrias do doutorado.

Aos amigos Czar, Severino e Luizinho pelo incentivo, confiança e apoio com revisão de artigos e de traduções.

Ao Alexandre que igualmente assumiu o desafio de realizar o doutorado em paralelo com o trabalho.

Aos amigos e amigas da CTDE Brenda Rocco, Carlos Ditadi, Claudia Lacombe, Daniel Flores, Humberto Innarelli, João Lima, Luiz Sayão, Margareth da Silva, Neire Martins, Rosely Rondinelli, Vanderlei Santos que muito ajudaram com discussões de conceitos e sugestões.

Aos amigos e amigas que acompanharam e torceram pelo sucesso da empreitada.

E a Ilma, Letícia e Luísa que com carinho e amor foram a fonte de energia para concluir este trabalho.

過去と未来の錯綜する
野生の島

月島 玲児

RESUMO

O contínuo aumento do volume da informação nas organizações é decorrente do crescimento das atividades desenvolvidas pela sociedade em geral e das facilidades proporcionadas pela tecnologia da informação em constante evolução. Paradoxalmente, esse rápido desenvolvimento de novas tecnologias que facilitam a criação de conteúdos ameaça a sua preservação, tornando-os tecnologicamente obsoletos em curto espaço de tempo. Nesse contexto da mutabilidade da tecnologia e do grande volume de objetos digitais, o objetivo deste trabalho é a concepção e a proposição de um *framework* de preservação digital de longo prazo, visando à preservação do conhecimento. Com esse propósito, recorreu-se aos fundamentos, às técnicas e às metodologias da Engenharia do Conhecimento, Gestão do Conhecimento, Arquivologia, Biblioteconomia e Computação, como o processo de avaliação desenvolvido na Arquivologia e no modelo de referência OAIS. A validação do *framework* foi realizada por meio de entrevistas semiabertas a especialistas em Computação, Gestão do Conhecimento e Preservação de Documentos Digitais. Todos os especialistas consideraram o *framework* proposto aplicável e reconheceram a inviabilidade de guardar todo o conteúdo digital, portanto, confirmando a importância de um processo de avaliação e seleção do objeto a ser preservado. Ajustes foram introduzidos no *framework* com base nas sugestões apresentadas pelos entrevistados. Constatou-se que a seleção de objetos digitais visando à preservação do conhecimento poderá ser realizada com segurança se os critérios de avaliação forem claramente estabelecidos e que a preservação do conhecimento codificado em objetos digitais é possível, desde que haja o alinhamento entre o objeto e a tecnologia que possibilita o seu acesso. Consequentemente, a gestão da dependência tecnológica dos objetos digitais é parte integrante do processo de preservação digital.

Palavras-chave: Engenharia e Gestão do Conhecimento; Preservação Digital; Objetos Digitais; Avaliação Documental.

ABSTRACT

The continued growth of information volumes in organizations results from the growth of the activities developed by the society in general and of the facilities offered by the constant evolution of information technology. The fast development of new technologies that facilitate the creation of content threatens its preservation and makes it technologically obsolete in a short time. In the context of technological mutability and huge volumes of digital objects, the objectives of this work are the design and the proposal of a framework for long-term digital preservation, aiming the preservation of knowledge. For this purpose, we refer to the fundamentals, techniques and methodologies of knowledge engineering, knowledge management, archival science, library science and computing, such as the appraisal process developed in the archival science and the OAIS reference model. The validation of the framework was realized through semi-structured interviews with specialists in computing, knowledge management and preservation of digital documents. All specialists considered the proposed framework applicable and recognized the impossibility of keeping whole digital content. This confirms the importance of the process of appraisal and selection of objects that will be preserved. Adjustments were introduced in the framework based on the suggestions given by interviewees. The conclusion was that the selection of digital objects for the preservation of knowledge can be safely performed if the appraisal criteria are clearly stated. The preservation of the knowledge encoded in digital objects is possible if the object and the technology that enables your access are aligned. Therefore, the management of digital objects technological dependence is an important component of the digital preservation process.

Keywords: Knowledge Engineering and Management; Digital Preservation; Digital Objects; Document Appraisal

Lista de Figuras

Figura 1 –V epistemológico de Gowin	32
Figura 2 – Formas de incorporar conhecimento de Fleck	46
Figura 3 – <i>Framework</i> para categorias de conhecimento	47
Figura 4 – Modelo de gestão do conhecimento-informação	48
Figura 5 – Balanceamento do conhecimento-informação	48
Figura 6 – Dimensões da criação do conhecimento de Nonaka e Takeuchi	50
Figura 7 – Tipos de conhecimento no nível cognitivo	51
Figura 8 – Três mundos de Popper	53
Figura 9 – Grupo 1 de entidades do FRBR	55
Figura 10 – Grupo 2 de entidades do FRBR	56
Figura 11 – Grupo 3 de entidades do FRBR	57
Figura 12 – Níveis das entidades do grupo 1 FRBR	58
Figura 13 – Pirâmide DIKW	60
Figura 14 – Níveis de informação e Processos de elaboração Informacional ...	61
Figura 15 – Funil da sabedoria	64
Figura 16 – Pirâmide DIK	66
Figura 17 – Ciclo Conhecimento-Informação.....	67
Figura 18 – Espiral do conhecimento	69
Figura 19 – Mapa mental da classificação do conhecimento	71
Figura 20 – Exemplo de Mapa Conceitual.....	75
Figura 21 – Diagrama <i>Record Continuum</i>	82
Figura 22 – Modelo de Ciclo de vida de Documentos Arquivísticos	90
Figura 23 – Técnicas de Preservação Digital	93
Figura 24 – Modelo de ciclo de vida de curadoria digital.....	97
Figura 25 – Correspondência entre níveis de preservação e objetos digitais ..	102
Figura 26 – Modelo de Referência OAIS	106
Figura 27 – Objeto digital no OAIS.....	108
Figura 28 – Modelo de Dados do PREMIS	109
Figura 29 – Estrutura do <i>Open Document Format</i>	114
Figura 30 – Exemplo de um arquivo ODF descompactado.	115
Figura 31 – Correspondência entre a estrutura ODF com o OAIS.....	117
Figura 32 – Principais conceitos adotados no <i>framework</i>	120
Figura 33 – Quadrantes de Tabelas de Decisão	121
Figura 34 – Espiral de desenvolvimento da ontologia DeTec.....	122
Figura 35 - Processos básicos de preservação do conhecimento	126
Figura 36 – Infográfico do quadro geral do <i>framework</i>	127
Figura 37 – <i>Framework</i> de Preservação de Objetos Digitais.....	128
Figura 38 – Estratégia do "peneiramento sucessivo"	133
Figura 39 – Componente Temporalidade.....	136
Figura 40 – Critérios de avaliação	137
Figura 41 – Componente Avaliação & Reavaliação	139
Figura 42- Classes da ontologia Detec.....	142
Figura 43 – Ontologia DeTec visualizado pelo OWLViz.....	143

Figura 44 – Ontologia DeTec	144
Figura 45 – Componente Dependência Tecnológica	147
Figura 46 – Componentes do grupo Descrição & Representação.....	148
Figura 47 – Pacote de Informação de Conteúdo	151
Figura 48 – Componente Atualização	153
Figura 49 – Componente Armazenamento	154
Figura 50 – <i>Framework</i> Visão Geral Resumida	156
Figura 51 – <i>Framework</i> de Preservação de Objetos Digitais com ajuste.....	160
Figura 52 – Processo do Componente Avaliação & Reavaliação Ajustado.....	162
Figura 53 – Componente Verificação de Cópia Ajustado	164
Figura 54 – Conhecimento e Objetos Digitais e Três Mundos de Popper	166
Figura 55 – Objeto de Dado, Informação e Conhecimento	167
Figura 56 – V epistemológico de Gowin completo	168
Figura 57 – Visão completa detalhada do <i>Framework</i>	201
Figura 58 – Visão completa detalhada do <i>Framework</i> Ajustado	202

Lista de Quadros

Quadro 1 – Motivos do esquecimento do conhecimento de objetos digitais	37
Quadro 2 – Espécies e formas de conhecimento.....	42
Quadro 3 – Tipos de conhecimento de Musgrave.....	43
Quadro 4 – Tipos de conhecimento de Lundvall	43
Quadro 5 – Tipos de conhecimento de Collins	44
Quadro 6 – Tipos de conhecimento segundo Millar, Demaid e Quintas.....	44
Quadro 7 – Tipos de conhecimento de Blackler	44
Quadro 8 – Formas de conhecimento de Fleck	45
Quadro 9 – Categorias de conhecimento de Blumentritt e Johnston.....	46
Quadro 10 – Quatro tipos de conhecimento de Lam.....	51
Quadro 11 – Dado, Informação e Conhecimento.....	62
Quadro 12 – Contraparte negativa do DIKW	65
Quadro 13 – Questões para análise do valor probatório	86
Quadro 14 – Fatores para teste de valor informativo	86
Quadro 15 – Decisão sobre o destino da informação.....	88
Quadro 16 – Técnicas de Preservação Digital	93
Quadro 17 – Diretrizes para o Preservador	95
Quadro 18 – Diretrizes para o Produtor	95
Quadro 19 – Categorias de propriedades significativas	104
Quadro 20 – Níveis de Interoperabilidade Conceitual.....	112
Quadro 21 – Correspondência entre a estrutura ODF e OAIS	118
Quadro 22 – Aplicação da tabela de decisão no processo de avaliação	138
Quadro 23 – Resultado da tabela de decisão do processo de avaliação	138
Quadro 24 – Qualificadores do elemento Preservação do e-PMG.....	149

Lista de Abreviaturas e Siglas

ABNT	Associação Brasileira de Normas Técnicas
ASCII	American Standard Code for Information Interchange
CCSDS	Consultative Committee for Space Data Systems
CMS	Content Management System
DCC	Digital Curation Centre
DCMI	Dublin Core Metadata Initiative
DeTec	Ontologia de Dependência Tecnológica
DIKW	Data-Information-Knowledge-Wisdom
EC	Engenharia do Conhecimento
EGC	Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento
FGDC	Federal Geographic Data Committee
FRBR	Functional Requirements for Bibliographic Records
GC	Gestão do Conhecimento
GCO	Gestão do Conhecimento Organizacional
GED	Gestão Eletrônica de Documentos
IDC	International Data Corporation
IEEE	Institute of Electrical and Electronics Engineers
IES	Instituição de Ensino Superior
IFLA	International Federation of Library Associations and Institutions
IHMC	Institute for Human and Machine Cognition
InSPECT	Investigating Significant Properties of Electronic Content
LOM	Learning Object Metadata
MER	Modelo de Entidade-Relacionamento
METS	Metadata Encoding & Transmission Standard
OAIS	Open archival information system
OASIS	Organization for the Advancement of Structured Information Standards
OCLC	Online Computer Library Center
ODF	Open Document Format
ODIN	Objetos Digitais de Identificação Negativa
OWL	Web Ontology Language
PREMIS	Preservation Metadata: Implementation Strategies
RLG	Research Library Group
SAAI	Sistema Aberto de Arquivamento de Informação
TIC	Tecnologia da Informação e Comunicação

UNESCO

Organização das Nações Unidas para a Educação, a
Ciência e a Cultura

W3C

World Wide Web Consortium

XML

eXtensible Markup Language

SUMÁRIO

1	Introdução.....	25
1.1	IDENTIFICAÇÃO DO PROBLEMA	26
1.2	QUESTÃO DE PESQUISA.....	31
1.3	OBJETIVOS ESPECÍFICOS	32
1.4	V EPISTEMOLÓGICO DE GOWIN	32
1.5	JUSTIFICATIVA	33
1.6	ADERÊNCIA AO EGC	37
1.7	LIMITAÇÕES DA PESQUISA.....	38
1.8	ETAPAS DO DESENVOLVIMENTO DO TRABALHO	39
2	Fundamentos Teóricos	40
2.1	O CONHECIMENTO	40
2.1.1	A Tese dos Três Mundos de Popper	52
2.1.2	FRBR.....	54
2.1.3	Dado – Informação – Conhecimento	58
2.2	GESTÃO DO CONHECIMENTO	68
2.2.1	Conhecimento e GC: Síntese dos aspectos de interesse.....	70
2.3	ENGENHARIA DO CONHECIMENTO.....	73
2.3.1	Ontologia	73
2.3.2	Mapas Conceituais	74
2.4	ARQUIVOLOGIA	76
2.4.1	O repensar da ciência arquivística.....	77
2.4.2	Avaliação	84
2.4.2.1	Operação Metodológica de Silva e Ribeiro	88
2.4.3	Conhecimento da Arquivística na preservação do conhecimento	89
2.5	PRESERVAÇÃO DIGITAL.....	92
2.5.1	Abordagens e técnicas de Preservação Digital.....	92
2.5.2	Abordagem Sociotécnica de Preservação	94
2.6	CURADORIA DIGITAL.....	96
2.7	OBJETO DIGITAL	99
2.7.1	As classes de objetos digitais de Thibodeau	100
2.7.2	Propriedades significativas dos objetos digitais.....	103
2.8	MODELO DE REFERÊNCIA OAIS	104
2.9	DICIONÁRIO DE DADOS PREMIS	109
2.10	INTEROPERABILIDADE	111
2.10.1	Interoperabilidade Temporal.....	112
2.11	OPEN DOCUMENT FORMAT – ODF	113
2.11.1	Representação e descrição no ODF e OAIS	116
3	Metodologia e Alinhamento Teórico.....	119

3.1	AVALIAÇÃO E REAVALIAÇÃO	121
3.2	ONTOLOGIA DE DEPENDÊNCIA TECNOLÓGICA	122
3.3	VERIFICAÇÃO E AJUSTES DO <i>FRAMEWORK</i>	123
4	Framework de Preservação de Objetos Digitais	126
4.1	DESCRIÇÃO DO FRAMEWORK.....	126
4.2	COMPONENTES DA RECEPÇÃO.....	129
4.2.1	Objeto Digital	129
4.2.2	Componentes identificação inicial e limpeza	130
4.2.3	Componente verificação de cópia	130
4.3	COMPONENTES DA SELEÇÃO	131
4.3.1	Componente Colaboração	133
4.3.2	Componente Temporalidade.....	134
4.3.3	Componente Avaliação e Reavaliação.....	136
4.4	COMPONENTES DAS AÇÕES DE PRESERVAÇÃO	139
4.4.1	Componente Dependência Tecnológica.....	140
4.4.1.1	Ontologia DeTec	141
4.4.1.2	Processo de verificação da Dependência Tecnológica.....	146
4.4.2	Componente Informação Descritiva.....	147
4.4.3	Componente Informação de Representação	150
4.4.4	Atualização	152
4.5	ARMAZENAMENTO	153
4.6	FRAMEWORK – VISÃO GERAL	155
5	Verificação e ajustes do Framework	157
5.1	ANÁLISE DAS RESPOSTAS	157
5.1.1	Aplicabilidade do framework nas organizações	157
5.1.2	O processo de seleção dos objetos digitais.....	157
5.1.3	Grupo ações de preservação.....	158
5.1.4	Identificação dos pontos fortes do framework.....	158
5.1.5	Identificação dos pontos para Melhorias	158
5.1.6	Grau de satisfação	159
5.2	AJUSTES REALIZADOS NO FRAMEWOK.....	159
5.2.1	Gestor do framework.....	159
5.2.2	Parcialidade no processo de seleção	161
5.2.3	Objetos assinados com certificado digital	161
5.2.4	Avaliação com resultado falso negativo	161
5.2.5	Documentos idênticos como documentos distintos.....	163
5.2.6	Ajuste do diagrama geral do framework	164
6	Conclusões.....	165
6.1	TRABALHOS FUTUROS.....	170

Referências.....	172
Apêndice A – Questões-chave das entrevistas.....	190
Apêndice B – Perfil dos entrevistados	191
Apêndice C – Síntese das respostas.....	193
Apêndice D – visão geral do Framework	201
Apêndice E – Visão Geral do framework Ajustado	202

1 INTRODUÇÃO

Fixar a informação em algum tipo de suporte acompanha a história da evolução do conhecimento. Os primeiros registros de informação, as pinturas rupestres, são datados por paleontólogos em 40.000 a.C.

O suporte da informação migrou para outros objetos, como pedras, placas de barro, metais, pergaminho, papiro. A adoção posterior do papel como suporte de informação e a invenção da imprensa permitiu a reprodução em escala de livros e documentos, o que leva ao raciocínio de que, provavelmente, Gutenberg é um dos responsáveis pela explosão informacional.

O desenvolvimento e popularização da tecnologia da informação e comunicação (TIC) intensificou a produção e disseminação da informação, acarretando profundas transformações na sociedade. Esse crescimento vertiginoso da produção da informação foi comprovado em projetos e estudos, como os desenvolvidos pela Universidade da Califórnia (LYMAN; VARIAN, 2000) e pela *International Data Corporation* (IDC) (GANTZ; REINSEL, 2010; GANTZ *et al.*, 2007).

O rápido e constante desenvolvimento da TIC permite a crescente produção de novas informações e conhecimentos, porém, paradoxalmente, esse rápido desenvolvimento é, talvez, o principal motivo do perecimento das informações armazenadas na forma digital. As mudanças na tecnologia tornam os registros criados e armazenados em tecnologias anteriores inacessíveis. Embora as cadeias de *bit* possam permanecer gravadas em suportes magnéticos ou óticos, a falta de uma tecnologia que consiga decodificá-las em sinais inteligíveis leva à perda total do conteúdo.

A necessidade de preservação sempre existiu, mas a preservação digital tem diferenças fundamentais em relação à preservação dos registros em papel. Enquanto no registro em papel a preservação do suporte geralmente resulte na preservação do conteúdo, podendo eventuais perdas ser parciais, no caso do registro digital, a preservação do suporte não é suficiente e a perda, se houver, em regra, é total. Pesquisadores podem reconstituir partes da história a partir de fragmentos de documentos criados há centenas de anos, enquanto fitas magnéticas gravadas há alguns anos podem ter o conteúdo totalmente inacessível.

A maioria dos produtos e serviços no mercado atual não existia há cinco anos e não tem métodos comprovados para garantir que as informações continuarão a existir e que serão acessadas com as

tecnologias disponíveis (CHEN, 2001). É difícil imaginar a era da informação sem o acesso ilimitado dos dados digitais que é o seu alicerce (BERMAN, 2008).

Na segunda metade do século passado, o crescimento do volume informacional estimulou as pesquisas que permitissem a busca e localização da informação desejada, resultando em significativos avanços nas técnicas e tecnologias de recuperação da informação. A localização da informação desejada na Internet seria muito mais morosa e complexa sem um mecanismo de recuperação (ou mecanismo de busca). O desafio atual é o desenvolvimento de técnicas e tecnologias para preservar o conteúdo armazenado em registros digitais.

Este trabalho apresenta a proposta de um *framework* para preservação digital, numa visão interdisciplinar, que, para o seu desenvolvimento, buscou o constante diálogo entre a Engenharia do Conhecimento, a Gestão do Conhecimento, a Arquivologia, a Ciência da Computação e a Ciência da Informação.

1.1 IDENTIFICAÇÃO DO PROBLEMA

A adoção do papel como suporte tornou o registro de informações e conhecimento corriqueiro. O papel possibilitou o registro na forma de texto, fotos, desenhos, diagramas e gráficos, resultando em registros usualmente denominados documentos, que tem como característica serem estáticas e estáveis; documento, nesse caso, é uma unidade de registro de informações em qualquer suporte ou formato (ARQUIVO NACIONAL, 2005).

Uma propriedade crítica de documentos por meio de gerações tem sido a sua fixidez, isto é, a habilidade de marcar a superfície do suporte de forma relativamente estável para possibilitar que pessoas, ao longo do tempo, possam visualizar as mesmas imagens e, então, terem acesso ao mesmo significado ou intenção comunicativa (LEVY, 1994).

A eletrônica introduziu novas formas de registro com a incorporação do som e da imagem em movimento (filmes), e os registros de conhecimento são armazenados em formatos digitais. Com isso, os usuários ganham novas facilidades para acessar, criar e alterar informações e conhecimentos registrados. A interatividade das pessoas com os registros de conhecimento foi uma importante característica introduzida pela TIC. Atualmente, os registros médicos, dados financeiros, e outras informações pessoais e profissionais estão na forma digital (BERMAN, 2008).

A transição do papel para registros digitais dos documentos trouxe inúmeras vantagens, como a facilidade na criação, disseminação e interatividade. Porém, a fixidez que é a qualidade do documento manter a sua forma fixa e o conteúdo estável (INTERPARES PROJECT, 2007) é uma fragilidade dos registros digitais.

Paradoxalmente, a evolução da TIC que, por um lado, introduz constantes melhorias e facilidade nas interações humana-máquina, por outro lado, torna as tecnologias existentes obsoletas em pouco tempo, dificultando ou impedindo o acesso aos registros criados nas tecnologias antecessoras.

Essa obsolescência tecnológica ocorre tanto no *hardware* com dispositivos cada vez menores, mais rápidos e com maior capacidade de armazenamento, assim como no *software*, que evolui para novos sistemas operacionais, aplicativos e formatos de arquivos.

Infelizmente, não existe garantia de preservação e acessibilidade da informação digital gerada no contexto do rápido avanço tecnológico (CHEN, 2001) e os *softwares* disponíveis atualmente não incluem boas ferramentas para salvar os originais digitais no contexto da rápida obsolescência do *hardware* e *software* (GLADNEY, 2007), no entanto, usuários desejam armazenar e usar seus dados por períodos que vão de dias a décadas (BERMAN, 2008).

Constantopoulos e Dritsou (2007) classificam os riscos aos quais os ativos informacionais estão sujeitos. O risco técnico é decorrente da evolução técnica do *hardware* e *software*, e o risco físico, de danos na mídia de armazenamento e incidentes catastróficos.

Seguindo raciocínio semelhante, a taxonomia de preservação apresentado por Graham (1994) apresenta três tipos de preservação: a preservação de mídia; a preservação tecnológica e a preservação intelectual. A preservação de mídia endereça o risco físico; a preservação tecnológica tem foco no risco técnico; e a preservação intelectual endereça a integridade e autenticidade da informação originalmente gravada (GRAHAM, 1994).

Encaminhar os três tipos de riscos já é um problema complexo, todavia, considerando os outros fatores, como o grande volume de dados e informações, e a dificuldade em selecionar os objetos que devem ser preservados, o problema da preservação digital torna-se um desafio de altíssima complexidade.

A facilidade em usar, reusar, alterar e criar registros acelerou o crescimento do volume de novas informações. Praticamente todos os registros de informações, na atualidade, nascem digitalmente, embora a versão física seja exigida em algumas situações.

Em 2000, o projeto “*How Much Information?*” concluiu que o mundo produzia entre 1 e 2 exabytes de informação por ano e que somente 0,003% seria de documentos impressos (LYMAN; VARIAN, 2003).

Estudos realizados pelo IDC apontaram que, em 2006, a quantidade de informação digital criada, capturada e replicada foi de 161 exabytes ou 161 bilhões de gigabytes, e estimou que, entre 2006 e 2010, o incremento anual de informação no universo digital passaria de 161 para 988 exabytes, num crescimento superior a seis vezes (GANTZ *et al.*, 2007). Em 2010, a estimativa atualizada previu que, em 2020, o Universo Digital será 44 vezes maior do que foi em 2009, passando de 0,8 para 35 Zettabytes¹ (GANTZ; REINSEL, 2010).

Outro exemplo de volume do conteúdo digital foi divulgado pelo popular site de publicação de vídeos da Internet, YouTube (www.youtube.com). Em seu blog oficial, anunciou, em novembro de 2010, que o volume de publicação no site atingira 35 horas de vídeos por minuto, aproximadamente, 2.100 horas de vídeo a cada 60 minutos, 50.400 horas por dia. Comparado a filmes de longa-metragem de 120 minutos de duração, seria equivalente ao lançamento semanal de mais de 176 mil filmes². Numa evidente tendência de crescimento, em meados de 2007, o volume era de 6 horas de vídeo por minuto, cresceu para 15 horas em janeiro de 2009, para 20 horas em maio de 2009³ e, finalmente, 35 horas de vídeo por minuto em novembro de 2010.

Com o objetivo de gerir o grande volume de registros informacionais das instituições, diferentes vertentes resultaram em diferentes pontos de vista, propostas metodológicas e técnicas com diferentes denominações, como gestão de dados, Gestão da Informação, gestão de conteúdo e gestão de documento.

A proliferação de propostas teóricas, técnicas e tecnológicas evidencia a importância e a própria dificuldade na sua realização. A gestão, a organização, o acesso e a preservação dos dados digitais são grandes desafios na era da informação (BERMAN, 2008).

A gestão e a preservação são dois assuntos altamente relacionados. A falha em um afeta, diretamente, a realização do outro. Partindo da gestão, é possível concluir que um processo de gestão

¹ Zettabyte é igual a 1 trilhão de gigabytes.

² <http://youtube-global.blogspot.com/2010/11/great-scott-over-35-hours-of-video.html>

³ http://youtube-global.blogspot.com/2009/05/zoinks-20-hours-of-video-uploaded-every_20.html

adequado de registros de informação permite a sua preservação, por outro lado, sem um processo de preservação o próprio processo de gestão não se efetiva. Entre os dois assuntos, há pontos comuns e específicos. Pode-se dizer que são interdependentes – a realização de um depende do outro.

No tocante à gestão da Informação e do conhecimento, a considerável quantidade de publicação relacionada indica a preocupação e o reconhecimento da importância do assunto, no entanto, a realização adequada da gestão dos registros é ainda um desafio inalcançado pela significativa parte das instituições.

Como agravante, a mudança do objeto informacional físico para o digital ainda está em processo de absorção. O mundo corporativo e seus gerentes, acostumados a lidar com recursos físicos e tangíveis, esbarram em grandes desafios e dúvidas, como a forma gerir algo intangível como conhecimento ou o que deve ser feito para localizar, medir, usar e desenvolver o conhecimento (BLUMENTRITT; JOHNSTON, 1999).

A arquitetura de suporte dos dados digitais é a união coordenada de tecnologias da informação, sistemas, especialistas e organizações para possibilitar trabalho, recreação, pesquisa, educação e vida na era da informação (BERMAN, 2008).

As dificuldades em preservar o conteúdo digital podem deixar um período sem memória na história da humanidade, denominado por Kuny (1997) de *digital dark ages*.

A memória organizacional é a referência crucial para novas experiências sem a qual a aprendizagem é impossível, no entanto, o valor da memória organizacional é, hoje, subestimado pelas organizações que perdem parte da memória corporativa nas reorganizações, fenômeno batizado de amnésia coletiva ou síndrome do Alzheimer coletivo (ROMHARDT, 1997).

Para permanecerem competitivas, as organizações devem, efetivamente, criar, localizar, capturar e compartilhar seus conhecimentos e expertises, isso requer que o conhecimento organizacional seja explicitado e registrado para facilitar a distribuição e reúso (ZACK, 1999). Esse esforço de explicitação e de registro do conhecimento deverão, também, ser acompanhado de um processo de preservação desses registros armazenados a fim de evitar a volatilidade do conhecimento armazenado nesses registros digitais

O conhecimento explícito é o conhecimento que pode ser articulado na linguagem formal, afirmações gramaticais, especificações, manuais e é o conhecimento que pode ser transmitido entre as pessoas (NONAKA; TAKEUCHI, 1997, p XIII). O conhecimento explícito

recebe, também, a denominação de **conhecimento codificado** – objeto de pesquisa deste trabalho – por autores como Nonaka & Takeuchi (1997, p 327), Zack (1999), Blumentritt & Johnston (1999), Lundvall (1996; 2003), David & Foray (1995; 2002), Millar, Demaid e Quintas (1997) e Lam (1998).

O grande desafio do futuro não estará relacionado ao como armazenar a informação, estará, principalmente, relacionado à redução do custo de armazenamento, à redução do risco e à perda do conteúdo; e à extração de todo valor do conteúdo salvo (GANTZ; REINSEL, 2010).

A preservação é uma das funções primárias da Biblioteconomia e da Arquivologia, cada área com foco específico. A Biblioteconomia tem foco na preservação do acervo da biblioteca, como livros e revistas, e a Arquivologia com foco na preservação dos documentos de arquivo. A transição do papel para o digital de livros e documentos instigou a Biblioteconomia e a Arquivologia no estudo e desenvolvimento de técnicas, processos e tecnologias para a preservação digital.

Mais do que mudanças de técnicas e processos, uma questão complexa na transição da preservação dos objetos físicos para o digital é a necessidade da mudança de paradigma e entendimento do que é preservar. O paradigma e as definições pré-digitais não são úteis para aplicação na preservação digital (HARVEY, 2005, p. 7). O foco da preservação deve ser na preservação do conhecimento e não na preservação de itens individuais (CLOONAN, 1993; HARVEY, 2005).

Considerando os dados apresentados no estudo do Universo Digital (GANTZ; REINSEL, 2010; GANTZ *et al.*, 2007), grande parte do conhecimento codificado está armazenado em registros digitais na forma de documentos, especificações de produtos e serviços, normas, programas de computadores, modelo de dados, ontologia, projetos de engenharia, manuais de usuários, etc.

A preservação dos objetos digitais é também um assunto crítico para a Gestão do Conhecimento organizacional (GCO), pois a dificuldade em preservar os objetos digitais volatiliza o conhecimento neles armazenados. Métodos, processos e ferramentas que deem suporte na preservação do conhecimento codificado devem exercer importante papel para a GCO.

Embora com objetos e focos específicos, o arcabouço de conhecimento desenvolvido na Biblioteconomia e Arquivologia trazem importantes contribuições para a preservação do conhecimento codificado, a exemplo do processo de avaliação desenvolvido na Arquivologia.

Avaliação é o ato de julgar o valor primário e secundário de documentos arquivísticos e estabelecer o período de tempo durante o qual eles mantêm esse valor (COUTURE, 2005). É o processo de seleção de documentos arquivísticos para a preservação de longo prazo, tradicionalmente realizadas por pessoas, tornando-o trabalhoso, moroso e de alto custo.

A avaliação é uma tentativa de responder algumas das questões-chave dos projetos de preservação (MACEVICIUTE, 2012):

o que deveria ser preservado?

é possível preservar tudo?

é necessário preservar tudo?

como selecionar objetos dignos de preservação?

Se o processo de decidir o que pode ser descartado é complexo, a preservação da totalidade é também problemática. A preservação digital é imperfeita, com limitações tecnológicas e de fatores não técnicos (KASTELLEC, 2012).

1.2 QUESTÃO DE PESQUISA

Para a preservação do conhecimento registrado nos objetos digitais das organizações, um processo de seleção dos registros que devem ser preservados pode ser uma importante contribuição para a viabilidade do projeto de preservação.

No complexo cenário da preservação digital, este trabalho tem foco na seguinte pergunta:

Quais componentes são basilares em um *framework*⁴ de preservação de conhecimento codificado armazenados em objetos digitais, considerando o volume e a mutabilidade constante da tecnologia?

Neste sentido, este trabalho tem como Objetivo Geral:

Conceber um *framework de preservação do conhecimento codificado*, armazenado em registros digitais, alicerçado nos princípios, fundamentos e técnicas da engenharia e gestão do conhecimento, e no arcabouço de conhecimento desenvolvido nas iniciativas e nos projetos de preservação digital.

⁴ Neste trabalho, o termo “*framework*” é utilizado com o sentido de uma estrutura que contempla um conjunto de conceitos, componentes e métodos, necessários para atingir um objetivo.

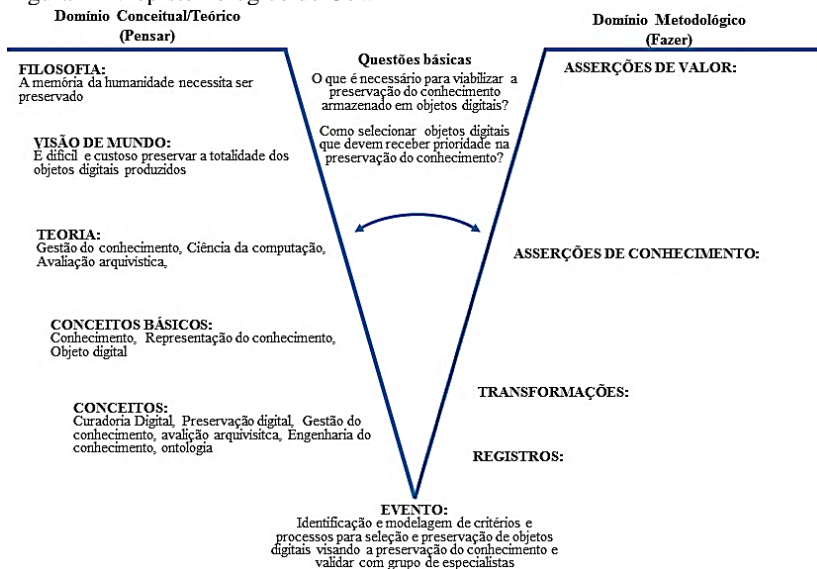
1.3 OBJETIVOS ESPECÍFICOS

- identificar princípios, métodos e técnicas de projetos que contribuam para a preservação do conhecimento;
- identificar objeto digital e suas propriedades;
- definir um método para avaliação de registros digitais com conhecimentos a serem preservados;
- elaborar um modelo conceitual visando à *preservação do conhecimento codificado*, alicerçado na *Engenharia do Conhecimento* (EC).

1.4 V EPISTEMOLÓGICO DE GOWIN

A fim de propiciar o melhor entendimento, o objeto do estudo foi representado com o uso do “V” epistemológico de Gowin (Figura 1).

Figura 1 –V epistemológico de Gowin



Fonte: Elaborado pelo autor.

O V de Gowin permite a conexão entre eventos, fatos e conceitos, que pode ser observada na forma de um “Vê” no qual é visualizada a ligação entre evento na ponta do vê, os conceitos e sistemas conceituais no lado esquerdo, e os fatos do lado direito (MOREIRA, 2006).

1.5 JUSTIFICATIVA

A preservação da informação é um dos assuntos mais importantes para a história, cultura e economia da humanidade, assim como para o desenvolvimento da civilização (LEE *et al.*, 2002).

A rápida obsolescência da tecnologia (ameaça técnica), a deterioração da mídia (ameaça física), e fatores como a dificuldade e o descuido na gestão, e o alto volume dos ativos digitais sintetizam as dificuldades e a complexidade do processo de preservação dos registros digitais e, por conseguinte, a preservação do conhecimento codificado armazenados nesses registros.

Uma política da preservação recai nas seguintes classes: a) migração do conteúdo digital; b) emulação de tecnologia; c) preservação de tecnologia; d) adoção de padrões; e) retrocedimento de tecnologia; f) encapsulação; g) identificador permanente; h) transformação para uma forma não digital; e i) arqueologia digital (CONSTANTOPOULOS; DRITSOU, 2007).

O grande volume, a complexidade e o alto custo da preservação digital são obstáculos que podem inviabilizar a preservação da totalidade dos objetos. Preservar tudo não é justificável nem necessário, pois, além do conhecimento perder a validade, significativa parte das informações digitais criadas não possuem a importância que justifique sua guarda. Porém, situações específicas possam levar à decisão de guardar tudo.

No estudo *Digital Universe*, de março de 2007, a *International Data Corporation* (IDC) previu que, em 2007, pela primeira vez, a quantidade de informações digitais criadas deveria exceder a quantidade de espaço de armazenamento disponível, embora, anualmente, a indústria lance milhares de petabytes de nova capacidade de armazenamento (GANTZ *et al.*, 2007).

Na atualização do estudo, em 2010, a conclusão foi que a diferença entre as informações criadas e o espaço de armazenamento tende ao crescimento, e a estimativa é que, próximo a 2020, cerca de 60% dos petabytes criados não poderão ser armazenados (GANTZ; REINSEL, 2010).

Em síntese, a produção da capacidade de armazenamento não acompanha o mesmo ritmo da produção da informação digital, consequentemente, mesmo desejando, não é possível manter a totalidade dos dados digitais produzidos (BERMAN, 2008).

O mesmo estudo que apresenta a preocupante previsão do déficit de capacidade de armazenamento divulga uma alentadora informação que, no imenso universo digital, somente algo em torno de 25% são de objetos únicos, portanto, 75% são cópias (GANTZ; REINSEL, 2010).

Dois fatores contribuem para o alto volume de cópias. O primeiro é a facilidade de copiar os objetos digitais e, segundo, a arquitetura dos sistemas que obriga a replicação do mesmo registro. Um documento anexado a um *e-mail* enviado a três pessoas, por exemplo, gera, no mínimo, quatro cópias do mesmo documento: um na caixa de *e-mail* do remetente e três cópias na caixa dos destinatários, e, muito provavelmente, uma quinta cópia no microcomputador ou servidor do remetente.

Poucos estudos revelam qual o percentual de registros de informação e conhecimento são destinados para a preservação de longo prazo. No caso de documentos arquivísticos, de acordo com estudos da UNESCO, do volume total de documentos criados e recebidos por uma entidade do governo, possivelmente, somente uma pequena porção, algo como 2% a 5%, tem suficiente valor administrativo, legal ou para pesquisa que justifique a guarda permanente (RHOADS, 1989).

Big Data é um termo recentemente incorporado na TIC para designar problemas e soluções tecnológicas para tratar as coleções de dados no atual contexto de volume, velocidade e variedade (CARTER, 2011; XEXÉO, 2013). O volume refere-se à grande quantidade de dados produzidos; a velocidade refere-se à alta taxa de bytes por intervalo de tempo enviados aos sistemas; e variedade refere-se às diferentes formas de dados produzidos (XEXÉO, 2013).

No contexto do Big Data, é frequente a necessidade de escolher o que guardar e descartar (XEXÉO, 2013), o que corrobora com a preocupação da Curadoria e Preservação Digital, e justifica o propósito de contemplar um processo de avaliação e seleção no *framework* proposto neste trabalho.

No contexto da preservação do conhecimento, todavia, o objeto da preservação, a motivação e os parâmetros de seleção devem ter particularidades distintas. Porém, se a proporção dos objetos digitais que devem ser preservados apresentar alguma semelhança com a estimativa da UNESCO e do estudo do IDC, o volume de registros que deverão receber tratamento para a preservação permanente deve ser significativamente reduzido em relação ao total.

Na Arquivologia, o processo que determina se o documento deve ser preservado ou eliminado é a Avaliação. Theodore R. Schellenberg, considerado o pai da avaliação arquivística americana, associa a

definição do termo arquivo ao processo de avaliação: “O elemento seleção deve estar implícito na sua definição de arquivos, pois o maior problema do arquivista atual consiste em selecionar, da massa de documentos oficiais criados por instituições públicas ou privadas de todos os gêneros, os arquivos que se destinam à preservação permanente” (SCHELLENBERG, 2006, p. 40).

A importância da avaliação dos documentos públicos (na época, em papel) foi defendida por Schellenberg (2006, p. 179) considerando: a) a impossibilidade do provimento de espaço de armazenamento; b) a necessidade de pessoal para cuidar; c) o custo da manutenção dos papéis que iam além das posses da mais rica nação; e d) a necessidade de reduzir a quantidade de documentos para que seja útil à pesquisa, caso contrário, os pesquisadores se desorientam ante a enorme quantidade de papéis. Portanto, a avaliação de documentos é considerada fundamental na arquivística.

Uma característica da avaliação documental é observada por Schellenberg (2006, p. 226): a impossibilidade de estabelecer padrões exatos para a avaliação. De acordo com o autor, não sendo exatos ou precisos, não necessitariam ser aplicados com absoluta consistência. Deveriam ser aplicados com moderação e bom senso, e com base na análise total da documentação relativa ao assunto a que se referem os documentos avaliados.

O conhecimento desenvolvido na Arquivologia subsidia o desenvolvimento de um modelo de avaliação de objetos digitais visando à preservação do conhecimento neles armazenados.

Considerando o volume e a velocidade exigida no atual cenário, o processo de avaliação com a subjetividade e a alta intervenção humana, como preconizado por Schellenberg, torna-se impraticável. Para o pai da avaliação arquivística, o trabalho humano não poderia ser substituído ou reduzido.

Não há possibilidade de serem inventadas técnicas que reduzam o trabalho de decidir sobre os valores dos documentos a uma operação mecânica. Não há, tampouco, um processo barato e fácil para se descartar documentos a não ser que se decida pela destruição de tudo que haja sido criado, jogando-se, por assim dizer, tudo fora. [...] **As dificuldades na avaliação** de documentos recentes **são tão grandes** que não admira que alguns arquivistas, em dado momento, **tendessem**

a fechar os olhos e nada fazer

(SCHELLENBERG, 2006, p.180, grifo nosso).

No atual contexto tecnológico, legal e administrativo, bem diferente daquele em que Schellenberg produziu a obra considerada referência até hoje na Arquivologia, e, principalmente considerando as diferenças entre a preservação do documento arquivístico e a preservação do conhecimento codificado, é passível que se acredite na viabilidade do desenvolvimento de um modelo que venha a reduzir o trabalho humano no processo de seleção e preservação do conhecimento codificado armazenado em registros digitais.

A importância da Gestão do Conhecimento (GC) nas organizações é reconhecida. Atualmente, o conhecimento é o recurso-chave de uma organização e a sua gestão tornou-se uma atividade crucial para as modernas organizações (SCHREIBER *et al.*, 2000).

O processo de criação do conhecimento está ancorado no pressuposto que é na constante interação entre o conhecimento tácito (subjetivo) e o conhecimento explícito (objetivo), que o conhecimento é criado e expandido (NONAKA; TAKEUCHI, 1997). Para que essa interação seja possível, a preservação e o acesso ao conhecimento codificado (explícito) é de fundamental importância.

A proteção do conhecimento é uma permanente batalha contra o processo natural de esquecer coisas. Neste sentido, Romhardt (1997) identifica os motivos que leva ao esquecimento do conhecimento armazenado em objetos digitais, o qual ele denomina de conhecimento “eletrônico”.

Quadro 1 – Motivos do esquecimento do conhecimento de objetos digitais

		Motivo
Memória apagada		Perda de dados irreversível por: <ul style="list-style-type: none"> • vírus; • problemas de hardware; • falhas de sistema; • falhas na cópia de segurança; • ataque hackers.
Perda de acesso	Limitada	<ul style="list-style-type: none"> • perda de dados reversível; • sobrecarga limitada; • problemas de interface
	Permanente	<ul style="list-style-type: none"> • incompatibilidade permanente de sistemas; • sobrecarga permanente; • codificação errada.

Fonte: Adaptado de Romhardt (1997).

Não por acaso, o assunto preservação digital tem despertado interesse de pesquisadores da comunidade científica e empresarial na busca de solução para evitar a *digital dark age* (KUNY, 1997) ou a amnésia coletiva (ROMHARDT, 1997).

1.6 ADERÊNCIA AO EGC

O foco do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento (EGC) é o ensino, a pesquisa, o desenvolvimento e a implementação de métodos e técnicas para a promoção da criação, da codificação (engenharia), do gerenciamento (gestão) e da disseminação (mídia) do conhecimento⁵.

Nos últimos anos, a Arquivologia tem desenvolvido conhecimento teórico-metodológico para a gestão, organização e preservação dos documentos arquivísticos digitais. A arquivística tem foco em documentos arquivísticos – um subconjunto da massa documental gerado e recebido pelas organizações.

A aplicação dos conhecimentos desenvolvidos na Arquivologia com o objetivo de preservar o conhecimento codificado armazenado em objetos digitais, numa visão unidisciplinar, certamente traria resultados, mas espera-se que, com a abordagem interdisciplinar, agregando os

⁵ www.egc.ufsc.br

conhecimentos da Engenharia do Conhecimento e da gestão do conhecimento, sejam alcançados resultados melhores e inovadores.

A EC provê metodologias científicas para analisar e engenhar o conhecimento (SCHREIBER *et al.*, 2000). Na sua evolução, deslocou do paradigma da *transferência* do conhecimento humano para a visão da *modelagem* (STUDER; BENJAMINS; FENSEL, 1998).

Assumindo a premissa que a seleção dos objetos que realmente devem ser guardados permanentemente viabiliza a preservação digital, a engenharia poderá contribuir especialmente na identificação das tarefas intensivas em conhecimento e na explicitação, e modelagem do conhecimento necessário para o *framework* de preservação.

Portanto, este trabalho tem a abordagem interdisciplinar, desenvolvido no contexto da Gestão do Conhecimento, contando com métodos e técnicas da Engenharia do Conhecimento, e alicerçado no conhecimento de disciplinas como a Arquivologia, Biblioteconomia e Computação.

1.7 LIMITAÇÕES DA PESQUISA

A preservação do conhecimento pode receber diferentes abordagens e enfoques. Este trabalho aborda especificamente a preservação do conhecimento que já esteja codificado e armazenado em registros digitais.

A preservação do conhecimento tácito e o processo de conversão do conhecimento tácito para explícito e do explícito para o tácito não fazem parte do escopo deste trabalho.

A preservação digital é uma área de crescente interesse e atenção de importantes centros de pesquisas e instituições, como os arquivos nacionais e as bibliotecas nacionais.

A preservação digital hoje é uma ampla área pesquisa podendo ter diferentes abordagens e focos. As estratégias de preservação digital são várias: preservação de tecnologia, refrescamento, emulação, migração/conversão, encapsulamento (FERREIRA, 2006).

Durabilidade da mídia, formatos de arquivos, desenvolvimento e especificação de repositórios de objetos digitais, aplicação de certificados digitais, são, entre outros, assuntos de pesquisa na área.

Esta tese tem como objeto a modelagem de um processo de preservação do conhecimento armazenado em objetos digitais, partindo da premissa que a seleção dos objetos que devem ser preservados viabiliza a processo.

Outros assuntos relacionados à preservação digital, como a durabilidade da mídia, emulação de tecnologia, virtualização de tecnologia e uso de certificados digitais para assegurar a autenticidade não serão tratados com profundidade.

1.8 ETAPAS DO DESENVOLVIMENTO DO TRABALHO

Este trabalho foi desenvolvido com as seguintes etapas:

- 1) aprofundar os fundamentos teóricos, técnicas e metodologias que auxiliem o desenvolvimento do *framework* de preservação do conhecimento armazenado em objetos digitais;
- 2) conceber um método para avaliação de objetos digitais que devem, prioritariamente, ser objeto de preservação;
- 3) desenvolver um *framework* com componentes que possibilitem a preservação do conhecimento codificado, armazenado em objetos digitais;
- 4) Validar o *framework* com um grupo de especialistas;
- 5) Realizar ajustes no *framework*.

2 FUNDAMENTOS TEÓRICOS

Quando se pretende gerir algo, normalmente, a preocupação primária é a busca da clareza conceitual do objeto a ser gerido. No caso da GC, a conceituação do seu objeto –conhecimento – não é uma tarefa trivial. Estudos, reflexões e discussões sobre o que é conhecimento há muito acompanham a humanidade e as transformações da sociedade ocasionadas pelo desenvolvimento da ciência e da tecnologia só fez aumentar o interesse no assunto.

Para o desenvolvimento deste trabalho buscou-se referências em assuntos e disciplinas que abrange aspectos conceituais e modelos referentes ao objeto a ser preservado, ou seja, o conhecimento codificado armazenado em objetos digitais:

- 1) Conhecimento;
- 2) Gestão do Conhecimento;
- 3) Engenharia do Conhecimento;
- 4) Arquivologia;
- 5) Preservação digital;
- 6) Curadoria Digital;
- 7) Objeto Digital;
- 8) Modelo de Referência OAIS;
- 9) Dicionário de Dados PREMIS;
- 10) Interoperabilidade; e
- 11) Open Document Format (ODF).

2.1 O CONHECIMENTO

Uma das partes essenciais da filosofia ocupa-se com o estudo da natureza, fontes e limites do conhecimento – a Teoria do Conhecimento (BAZARIAN, 1980, p.31; KLEIN, 2005).

O conhecimento como objeto de estudo tem a filosofia como precursora. Atualmente, outras áreas e disciplinas têm o conhecimento como seu objeto de estudo, a exemplo da Gestão do Conhecimento e a Engenharia do Conhecimento.

O conhecimento pode ser abordado sob a ótica do processo ou do produto.

Como processo, podemos definir o conhecimento como sendo o reflexo e a reprodução de nossa mente.

Como produto desse processo temos os conhecimentos sensíveis e racionais: os nossos

conhecimento de matemática, de física, biologia, psicologia, sociologia, filosofia, etc., que são estudados pelas respectivas disciplinas (BAZARIAN, 1980, p.43, grifo do autor).

Nessa visão da filosofia apresentado por Bazarian (1980), a preservação do conhecimento codificado e armazenado em objetos digitais tem foco no conhecimento como produto, porém, na elaboração do modelo de preservação, o correto entendimento do que é conhecimento, isto é, o conhecimento como processo, torna-se fundamental.

Além do conhecimento como processo e como produto, os diferentes entendimentos e uso do termo com distintos conceitos, tornando-o polissêmico, permitem variadas visões de mundo, dificultando o entendimento e o diálogo.

Uma das fases críticas antes de embarcar em iniciativas de Gestão do Conhecimento é a compreensão do conceito do objeto do gerenciamento – o conhecimento – com objetivo de identificar as fontes de conhecimento na organização e responder questões fundamentais que orientarão a concepção dessa iniciativa (LEE; FOO; GOH, 2006). Um melhor entendimento das características peculiares desta commodity é requisito para o efetivo gerenciamento do conhecimento (BLUMENTRITT; JOHNSTON, 1999).

O conhecimento é classificado e categorizado de diferentes formas (grau, tipo, forma, fonte, resultado, etc.), dependendo da visão de mundo do autor, e, talvez, devido à complexidade em conceituar e definir “conhecimento”.

As diferentes formas de classificar o conhecimento contribuem na identificação das características presentes nos diferentes usos e conceitos do termo conhecimento, e, talvez, a identificação da essência do conceito de conhecimento.

Nas classificações do conhecimento, autores utilizam os termos tipo de conhecimento, forma de conhecimento, categoria de conhecimento e componente como sinônimos. Neste trabalho, optou-se por manter o termo utilizado pelos autores.

Quanto ao grau, entre o conhecimento falso e verdadeiro, de forma geral, pode-se identificar três graus: o vulgar, o científico e o filosófico. O conhecimento vulgar é ametódico, isolado, casual e superficial, o conhecimento científico é causal, metódico dos fatos e fenômenos, e estabelece relacionamentos em busca de uniformidades e leis que o regem, e o conhecimento filosófico procura conhecer as causas reais dos fenômenos, com profundidade, universalidade e maior

grau de generalização que o conhecimento científico (BAZARIAN, 1980).

São igualmente três as espécies: conhecimento sensível ou empírico, conhecimento racional ou abstrato e conhecimento intuitivo ou criativo (BAZARIAN, 1980).

Quadro 2 – Espécies e formas de conhecimento

Espécies de conhecimento	Formas de conhecimento	Obtenção
Sensível ou Empírico	Sensação; Percepção; Representação	Órgãos sensoriais – visão, audição, tato, olfato, paladar
Racional ou Abstrato	Conceitos; Juízos; Raciocínio	Razão ou Intelecto
Intuitivo ou Criativo	Intuição sensível; Intuição racional; Intuição heurística	Intuição

Fonte: Adaptado de Bazarian (1980).

Essas classificações mais amplas do conhecimento refletem a definição do que é a filosofia atual: “Atualmente, a filosofia é uma ciência que estuda as leis mais gerais do ser, do pensamento, do conhecimento e da ação. É uma concepção científica do mundo como um todo da qual se pode deduzir certa forma de conduta” (BAZARIAN, 1980, p.37).

Davenport e Prusak (2000) definem conhecimento como uma mescla fluída de experiência, valores, informações contextuais, percepção especializada que provê uma base estrutura para avaliação e incorporação de novas experiências e informações. Tem origem e aplicação na mente dos conhecedores e estão frequentemente embutidos em documentos, repositórios, rotinas organizacionais, processos, práticas e normas.

Para o propósito de gestão, as categorias tradicionais do conhecimento são imprecisas e de difícil operacionalização (BLUMENTRITT; JOHNSTON, 1999). Com esse propósito, Blumentritt e Johnston propõem um *framework for categories of knowledge*, agrupando em quatro principais tipos, em que as categorias criadas por outros autores são enquadradas.

Para a contextualização e melhor entendimento desse *framework*, torna-se importante revisitar os autores que serviram de fonte e amparo da proposta, são eles: Musgrave, Lundvall, Collins, Millar, Blacker e Fleck.

No contexto da Teoria do Conhecimento, Musgrave (1993) identifica três tipos de conhecimento, apresentados no Quadro 3.

Quadro 3 – Tipos de conhecimento de Musgrave

Tipos	
<i>knowing-of</i> ou <i>knowing by acquaintance</i>	Conhecimento de coisas ou objetos, como sei quem é meu vizinho, conheço a torre Eiffel, conheço o cheiro do abacaxi.
<i>knowing-how</i>	Conhecimento de como fazer coisas, como jogar xadrez ou andar de bicicleta.
<i>knowing-that</i> ou <i>propositional knowledge</i>	Conhecimento de declarações ou proposições.

Fonte: Baseado em Musgrave (1993, p.6).

Na perspectiva da economia da aprendizagem, Lundvall (1996, 2003) identifica quatro tipos de conhecimento (Quadro 4). Lundvall define a economia da aprendizagem como uma economia em que as mudanças são rápidas, conseqüentemente, as habilidades tornam-se obsoletas, demandando novas e em que as atividades intensivas em conhecimento crescem com maior rapidez do que outras atividades.

Quadro 4 – Tipos de conhecimento de Lundvall

Tipos	Características	
<i>Know-what</i>	É o conhecimento sobre fatos, equivale à informação.	Pode ser facilmente codificado e transferido. A aprendizagem pode ocorrer por meio de leitura.
<i>Know-why</i>	Conhecimento sobre princípios e leis, importante para o desenvolvimento tecnológico.	
<i>Know-how</i>	Refere-se a habilidades, isto é, a capacidade de fazer alguma coisa.	Não é facilmente codificado e transferido. A aprendizagem ocorre, principalmente, pela experiência prática e interação social.
<i>Know-who</i>	Envolve a capacidade de estabelecer relações com grupos especializados a fim de delinear as competências. Inclui informação sobre quem sabe o que e quem sabe fazer o quê.	

Fonte: Baseado em Lundvall (1996, 2003).

Collins (1993, 1995) identifica quatro tipos de conhecimento/habilidades: *symbol-type knowledge*, *embodied knowledge*, *embrained knowledge* e *encultured knowledge*.

Quadro 5 – Tipos de conhecimento de Collins

Tipos	Características
<i>Symbol-type knowledge</i>	Pode ser transferido sem perdas na forma codificada em livros, computadores, etc.
<i>Embodied knowledge</i>	Conhecimento contido no corpo, com correspondente estrutura, desenvolvimento muscular, por exemplo, jogar tênis. Não pode ser transferido por sinais.
<i>Embrained knowledge</i>	Conhecimento contido na matéria física do cérebro. Algumas habilidades cognitivas têm relação com a estrutura física do cérebro.
<i>Encultured knowledge</i>	Ligado a grupos sociais. Acompanha as mudanças da sociedade.

Fonte: Baseado em Collins (1993, 1995).

A inovação envolvendo empresas geograficamente dispersas e conhecimentos de diferentes disciplinas depende do compartilhamento de conhecimento, transferência de tecnologia e aprendizagem, formando uma rede transorganizacional, nesse contexto, Millar, Demaid e Quintas (1997) propõem uma categorização do conhecimento, fazendo conexão com a proposta de Collins e de Lundvall.

Quadro 6 – Tipos de conhecimento segundo Millar, Demaid e Quintas

Tipos	Lundvall	
<i>Catalogue Knowledge</i>	<i>Know-what</i>	Simbólico e facilmente transferível
<i>Explanatory Knowledge</i>	<i>Know-why</i>	
<i>Process Knowledge</i>	<i>Know-how</i>	Cultural e sensível ao contexto, aprendido por meio da socialização.
<i>Social Knowledge</i>	<i>Know-who</i>	
<i>Experiential Knowledge</i>	<i>What was</i>	

Fonte: Baseado em Millar, Demaid e Quintas (1997).

Com base nas categorias de conhecimento sugeridas por Collins (1993), a proposta de Blackler (1995) é composta de cinco categorias de conhecimento.

Quadro 7 – Tipos de conhecimento de Blackler

Tipos	Características
<i>Embrained knowledge</i>	Depende de habilidade cognitiva.
<i>Embodied knowledge</i>	Orientado à ação e somente parcialmente explícito.
<i>Encultured</i>	Refere-se ao processo de atingir o entendimento compartilhado.

<i>knowledge</i>	
<i>Embedded knowledge</i>	Conhecimento que reside nas rotinas sistêmicas. É analisado em termos de sistemas, a exemplo do relacionamento entre tecnologias, funções, procedimentos formais e rotinas.
<i>Encoded knowledge</i>	Transmitida por sinais e símbolos, como em livros e manuais.

Fonte: Baseado em Blackler (1995).

A última referência-base do *framework* de Blumentritt e Johnston é a proposta de James Fleck (1997) que identifica seis formas de conhecimento na organização (Quadro 8), o local em que cada forma de conhecimento está incorporada e a relação entre eles, como ilustrado na Figura 2.

Quadro 8 – Formas de conhecimento de Fleck

Formas	Características
<i>Formal knowledge</i>	Conhecimento formalizado e codificado, fisicamente incorporado em teorias, fórmulas, etc. normalmente na forma escrita ou diagramática como em livros, adquirida por meio de processos educacionais.
<i>Instrumentalities</i>	Conhecimento incorporado no uso de ferramentas e instrumentos. Requer outros componentes (<i>informal</i> , <i>tacit</i> e <i>contingent</i>) e é adquirido por meio da demonstração e da prática.
<i>Informal knowledge</i>	Conhecimento incorporado na interação verbal e, às vezes, na forma escrita. Apreendida por meio da interação em ambiente específico.
<i>Contingent knowledge</i>	Incorporado em contextos específicos, distribuído, aparentemente, em informações triviais, algumas vezes, na forma de dado, adquirida por aprendizagem imediata (<i>on-the-spot</i>).
<i>Tacit knowledge</i>	Incorporado em pessoas, consolidado na prática e na experiência, transmitido por aprendizado e treinamento.
<i>Meta-knowledge</i>	Conhecimento incorporado na organização como a cultura geral e as premissas filosóficas. Podem ser particulares da organização e também da sociedade. São adquiridos por meio da socialização.

Fonte: Baseado em Fleck (1997).

Recebe destaque no trabalho de Fleck (1997) o *contingent knowledge*, (conhecimento contingente), que é uma forma particular de conhecimento que está incorporado no contexto da atividade de trabalho que não tem sido explicitamente abordada, embora frequentemente incluída em outras formas de conhecimento, como o tácito.

Figura 2 – Formas de incorporar conhecimento de Fleck



Fonte: Adaptado de Fleck (1997).

O *contingent knowledge* raramente se concentra em uma pessoa. Geralmente, é distribuído, aparentemente trivial e altamente específico no domínio particular da aplicação. O processo de desenvolvimento de tecnologia tem, frequentemente, dependência crítica do *contingent knowledge*, incluindo: a familiaridade com as operações envolvidas e as idiossincrasias do equipamento existente; conhecimento do conjunto de contatos industriais e comerciais necessários para conclusão do projeto; a avaliação das habilidades e atitudes das pessoas envolvidas, e um entendimento do ambiente de trabalho e o clima das relações (FLECK, 1997).

A categorização de conhecimento de Blumentritt e Johnston (1999) abarca os esquemas propostos por Musgrave, Lundvall, Collins, Millar, Blacker e Fleck (Quadro 9). As categorias de conhecimento adotadas para o *framework* de Blumentritt e Johnston são:

Quadro 9 – Categorias de conhecimento de Blumentritt e Johnston

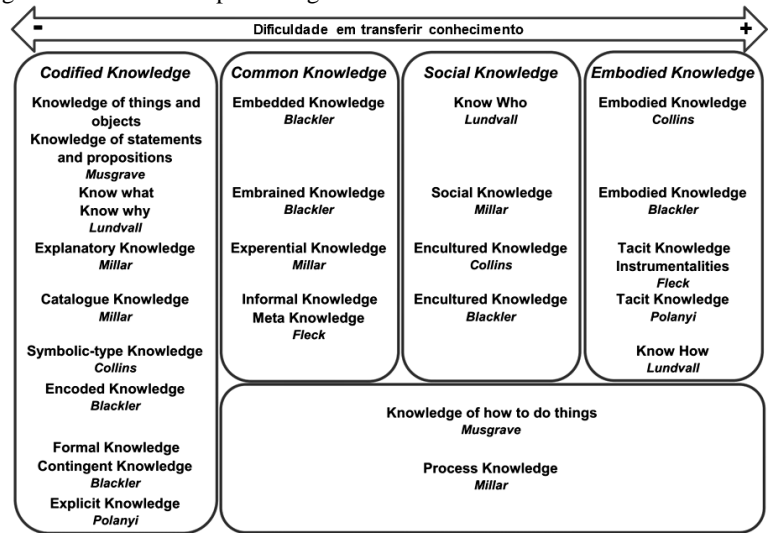
Categorias	Características
<i>Codified knowledge</i>	Equivalente à informação, explicitado por humanos por meio da escrita e outros meios de captura ou na forma de demonstração. Está em uma forma para transferência sem dificuldade.

<i>Common knowledge</i>	Conhecimento aceito como padrão mesmo sem ter sido formalmente explicitado, frequentemente na forma de práticas ou rotinas. Normalmente aprendida com o trabalho em um contexto particular.
<i>Social knowledge</i>	Conhecimento do relacionamento interpessoal e aspectos culturais, incluindo o conhecimento de a quem recorrer em determinada situação.
<i>Embodied knowledge</i>	Conhecimento acumulado por meio da experiência, histórico e habilidade da pessoa.

Fonte: Baseado em Blumentritt e Johnston (1999).

A relação entre as categorias é representada na Figura 3 que também identifica a grau de dificuldade da transferência de conhecimento, que é crescente da esquerda para direita.

Figura 3 – *Framework* para categorias de conhecimento

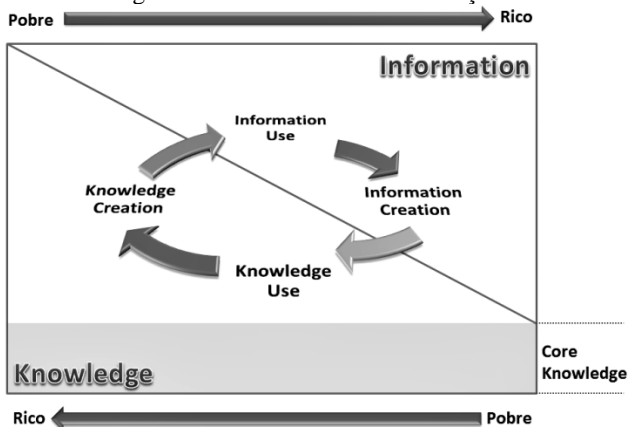


Fonte: Adaptado de Blumentritt e Jonhnston (1999).

Além de propor o *framework* para categorias de conhecimento, Blumentritt e Johnston (1999) apresentam uma proposta de modelo de gestão do conhecimento-informação (Figura 4). A assunção de que informação e conhecimento são distintos, e que o conhecimento requer um sistema inteligente é um dos aspectos centrais do modelo proposto, levando à definição do ciclo conhecimento-informação, no qual a

transformação do conhecimento em informação requer um processo de tradução.

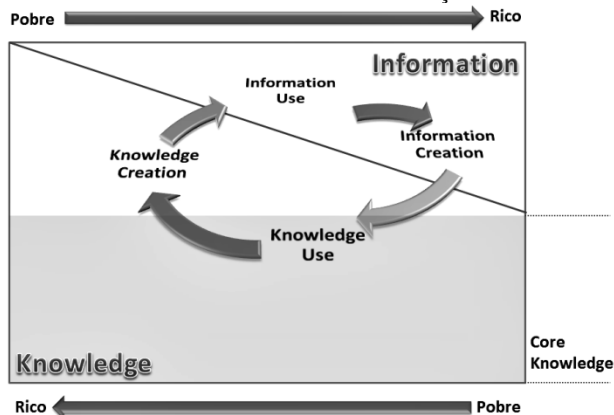
Figura 4 – Modelo de gestão do conhecimento-informação



Fonte: Adaptado de Blumentritt e Johnston (1999).

No modelo de Blumentritt e Johnston (1999), as novas tarefas que têm pouca informação necessitam de uma rica base de conhecimento (Figura 5). No entanto, as tarefas consolidadas, com rica base de informação, requerem menos conhecimento, principalmente o *core knowledge* (Figura 4).

Figura 5 – Balanceamento do conhecimento-informação



Fonte: Adaptado de Blumentritt e Johnston (1999).

O *core knowledge* é o conhecimento essencial para colocar a informação em ação. É construída a partir dos três tipos de conhecimento do lado direito do *framework* para categorização do conhecimento – *Common Knowledge*, *Social Knowledge* e *Embodied Knowledge* – e constitui no modelo o maior componente desse ativo intangível.

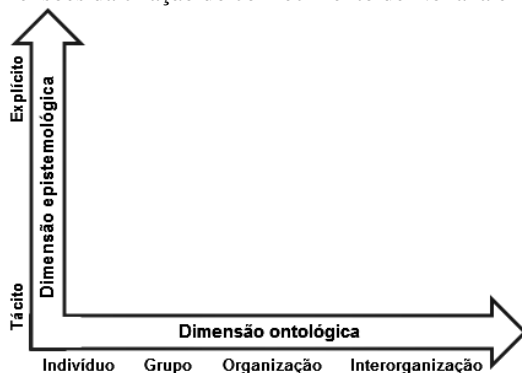
A partir do modelo de gestão do conhecimento-informação, os autores concluem que:

- uma pessoa possui maior *core knowledge* quando tem muitos contatos, experiência e habilidades;
- a relação entre conhecimento-informação pode ser distinto para as diferentes áreas de uma organização, permitindo o direcionamento de recursos. Muita informação para equipes com muito conhecimento pode ser desperdício de recurso, já a carência de informação em equipes com pouco conhecimento pode comprometer o sucesso da organização;
- durante a criação de informação na organização, o *core knowledge* para uso dessa informação deve ser identificado para permitir a distribuição dessa informação para usuários que tem o *core knowledge* adequado para criar novas informações.

Considerando que o conhecimento é multifacetado, complexo e dinâmico, e que existem muitos tipos de conhecimento que são relevantes para a organização, uma análise das características e dos atributos dos diferentes tipos de conhecimento permite explorar a natureza dos relacionamentos e construir uma tipologia do conhecimento organizacional (LAM, 1998).

No modelo de GC desenvolvido por Nonaka e Takeuchi (1997), são estabelecidas duas dimensões da criação do conhecimento – a epistemológica e a ontológica – em que a pedra fundamental da epistemologia é a distinção entre o conhecimento tácito e o explícito, e a dimensão ontológica é caracterizada pelo indivíduo que cria o conhecimento que é cristalizado como parte da rede de conhecimento da organização, podendo, na expansão, atravessar as fronteiras das organizações, tornando interorganizacionais (Figura 6).

Figura 6 – Dimensões da criação do conhecimento de Nonaka e Takeuchi



Fonte: Adaptado de Nonaka e Takeuchi (1997).

Lam (1998) utiliza a mesma epistemologia de Nonaka e Takeuchi, enquadrando na dimensão epistemológica o conhecimento explícito e tácito:

- o conhecimento explícito é a parte do conhecimento humano que pode ser especificado ou comunicado verbalmente ou na forma de símbolos, como documentos textuais, desenhos técnicos ou programas de computadores; e
- o conhecimento tácito é o conhecimento intuitivo, inarticulável, e não pode ser facilmente codificado e transferido.

Lam (1998), em sua análise, traz um elemento original ao relacionar os conhecimentos da dimensão epistemológica com a Tese dos Três Mundos⁶ de Popper (1999). Para a autora, o conhecimento explícito faz parte do mundo 3 de Popper, pois pode ser abstraído e armazenado no mundo objetivo, e entendido e compartilhado sem o detentor do saber. O conhecimento tácito faz parte do mundo 2 de Popper, em que o conhecimento não pode ser comunicado, entendido ou usado sem o detentor do saber.

A dimensão ontológica situa o local do conhecimento ou a entidade conhecedora, que pode ser no nível individual ou coletivo. Para essa dimensão ontológica, Lam (1998) observa que:

- o conhecimento individual é o conhecimento que reside na mente e nas habilidades incorporadas dos indivíduos, adquiridas por meio da educação formal e a prática, é discreto,

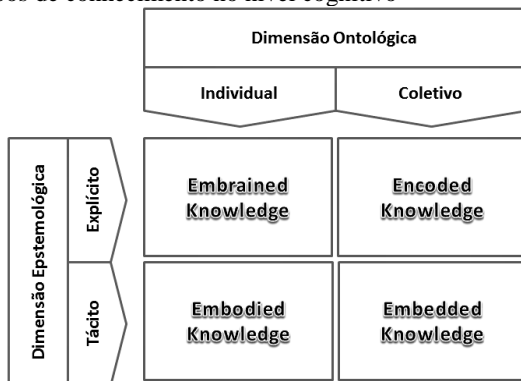
⁶ Tratado na seção 2.1.1 - A Tese dos Três Mundos de Popper.

independente e a autonomia na aplicação é sua característica-chave; e

- o conhecimento coletivo é a forma na qual o conhecimento é distribuído e compartilhado entre pessoas da organização. É o conhecimento acumulado armazenado em regras, procedimentos, rotinas e normas compartilhadas que guia o comportamento, a resolução de problemas e o padrão de interação entre as pessoas.

Dessas duas dimensões, emergem quatro formas de conhecimento organizacional: *embrained*, *embodied*, *encoded* e *embedded* (Figura 7), que tem base na distinção conceitual do conhecimento sugerida por Collins (1993) e adaptada por Blacker (1995), porém com foco na estrutura sociocognitiva do conhecimento, integrando as dimensões individuais e organizacionais como meio para relacionar as características do conhecimento (LAM, 1998).

Figura 7 – Tipos de conhecimento no nível cognitivo



Fonte: Adaptado de Lam (1998).

A definição para os quatro tipos de conhecimento, na visão de Lam (1998), é sintetizado no Quadro 10.

Quadro 10 – Quatro tipos de conhecimento de Lam

Formas de conhecimento	Dimensão	Características
<i>Embrained knowledge</i>	Individual-explicito	Dependente da habilidade do indivíduo, é formal, teórico ou abstrato, a exemplo do conhecimento científico, que tem foco no entendimento e saber racional de princípios universais ou leis da natureza. É adquirido,

		primariamente, com a educação formal e o treinamento. É o “aprender estudando”.
<i>Embodied knowledge</i>	Individual-Tácito	Orientado à ação, é o conhecimento prático. Tem um componente fortemente automático e voluntário, que, normalmente, não requer um processo consciente de tomada de decisão. É específico de contexto e pessoal. É o “aprender fazendo”.
<i>Embedded knowledge</i>	Coletivo-Tácito	É a forma coletiva do conhecimento tácito que reside nas rotinas organizacionais. É o conhecimento que reside no complexo relacionamento social que não pode ser totalmente articulado ou facilmente transferido. Está enraizado nas comunidades de prática. É o conhecimento baseado nas relações de pessoas, é contextual e disperso. É orgânico e dinâmico. É uma forma emergente de conhecimento capaz de suportar padrões complexos de interação na ausência de regras escritas.

Fonte: Baseado em Lam (1998).

Em um assunto complexo, uma aparente clareza conceitual pode, por vezes, introduzir interpretações equivocadas. Maior clareza de conceitos relacionados ao conhecimento pode ser apoiada em dois trabalhos, desenvolvidos em diferentes áreas do conhecimento, em distintos momentos e com objetivos distintos. O primeiro é a Tese dos Três Mundos desenvolvida pelo filósofo Karl Popper (1999), e o segundo, os Requisitos Funcionais para Registros Bibliográficos (Functional Requirements for Bibliographic Records - FRBR) — desenvolvido pela Federação Internacional de Associações de Bibliotecários (International Federation of Library Associations and Institutions – IFLA) (TILLET, 2003; LE BOEUF, 2007; INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2009).

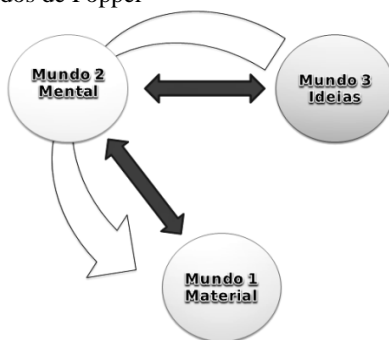
2.1.1 A Tese dos Três Mundos de Popper

A tese dos três mundos de Popper (1999) busca o pluralismo introduzindo o terceiro mundo para evitar o dualismo corpo-mente. Nessa visão, o mundo consiste em, pelo menos, três submundos ontologicamente distintos:

- o Mundo 1 é o mundo material, de objetos físicos ou dos estados materiais;
- o Mundo 2 é o mundo mental, de estados mentais ou de estados de consciência, ou, talvez, de disposições comportamentais para agir; e
- o Mundo 3, que é o mundo de conteúdos objetivos de pensamento, inteligíveis ou das ideias no sentido objetivo, o mundo de objetos de pensamentos possíveis, especialmente de pensamento científicos, poéticos e de obras de arte (POPPER, 1999).

Segundo Popper (1999, p. 152), o mundo 3 é “o mundo das teorias em si mesmas e de suas relações lógicas, dos argumentos em si mesmos, e das situações de problema em si mesmas”.

Figura 8 – Três mundos de Popper



Fonte: Baseado em Popper (1999).

Popper (1999) observa que essa filosofia pluralista tem como um dos problemas fundamentais a relação entre os três mundos em que os dois primeiros podem interagir e os dois últimos também, portanto, o mundo 2 – o mundo das experiências subjetivas ou pessoais – interage com o mundo 1 e o mundo 3, mas o mundo 1 e o mundo 3 podem interagir somente por meio do mundo 2. O mundo 2 é o mediador entre o mundo 1 e 3 (Figura 8).

O conhecimento codificado faz parte do conhecimento objetivo do mundo 3 de Popper (LAM, 1998) e, nesse ponto, torna-se importante buscar uma melhor compreensão dessa interpretação.

A tese de Popper envolve dois sentidos de conhecimento:

- 1) *conhecimento ou pensamento no sentido subjetivo* que é constituído de um estado de

espírito, ou de consciência ou de uma disposição para reagir; e 2) *conhecimento ou pensamento num sentido objetivo*, constituído de problemas, teorias e argumentos como tais. Neste sentido objetivo, o conhecimento é totalmente independente de qualquer alegação de conhecer que alguém faça; é também independente da crença ou disposição de qualquer pessoa para concordar; ou para afirmar, ou para agir. O conhecimento no sentido objetivo é *conhecimento sem conhecedor*; é *conhecimento sem sujeito que conheça* (POPPER, 1999, p. 110-111).

Portanto, a assunção de que o conhecimento objetivo de Popper é conhecimento sem conhecedor coincide com a propriedade do conhecimento codificado que não depende do sujeito conhecedor.

Mas não se deve confundir o conhecimento com o objeto no qual o conhecimento foi fixado. Não raramente os livros são indicados como exemplo do mundo 3 de Popper, porém um exemplar de um livro como objeto físico faz parte do mundo 1 de Popper, e o conhecimento nele registrado é o conhecimento objetivo que faz parte do mundo 3.

Um modelo que contribui para evitar a interpretação equivocada entre o objeto e o conteúdo é o FRBR desenvolvido pela IFLA.

2.1.2 FRBR

Os requisitos para registros bibliográficos foram estabelecidos como um modelo de referência que permite um entendimento comum de conceitos e denominações (LE BOEUF, 2007) para o universo bibliográfico.

O modelo da IFLA foi desenvolvido com o uso do modelo de entidade-relacionamento (MER) que é um método de modelagem proposto por Peter Chen (1976) para permitir a incorporação de informações semânticas em projetos de bancos de dados, utilizando técnica diagramática e com a intenção de ser usado como uma base para unificação de diferentes visões de dados.

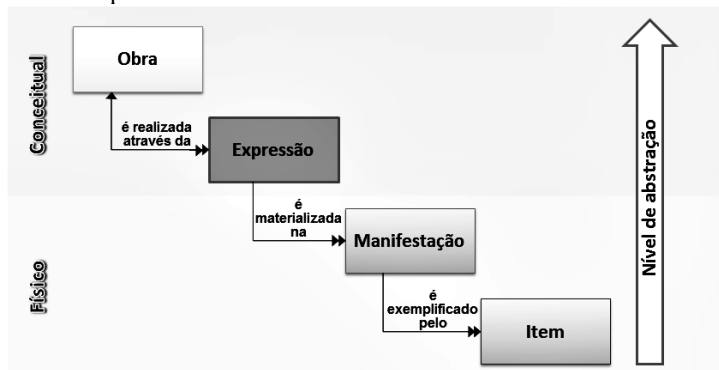
Os objetivos primários do estudo de requisitos funcionais para registros bibliográficos são dois: primeiro, o de prover uma estrutura claramente definida relacionada aos dados de registros bibliográficos, visando às necessidades dos usuários; e segundo, para recomendar um nível básico de funcionalidades de registros criados pelas entidades bibliográficas (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2009).

Os FRBR oferecem um vocabulário mais preciso para apoiar a construção de regras de catalogação, e uma estrutura e relacionamentos dos registros bibliográficos e de autoridade (TILLET, 2003).

O *Final Report* do IFLA *Study Group* estabelece os requerimentos funcionais para registros bibliográficos em três grupos de entidades (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2009).

O grupo 1 representa produtos do esforço intelectual ou artístico. Segundo Le Bouef (2007), é o núcleo dos FRBR. São quatro entidades – obra, expressão, manifestação e item – que pertencem ao documento em si, isto é, os “elementos” que são catalogados – Figura 9: **Obra** é uma criação intelectual ou artística distinta; **Expressão** é a realização intelectual ou artística de uma *obra*; **Manifestação** é a materialização de uma *expressão* de uma *obra*; e **Item** é um exemplar de uma *manifestação*.

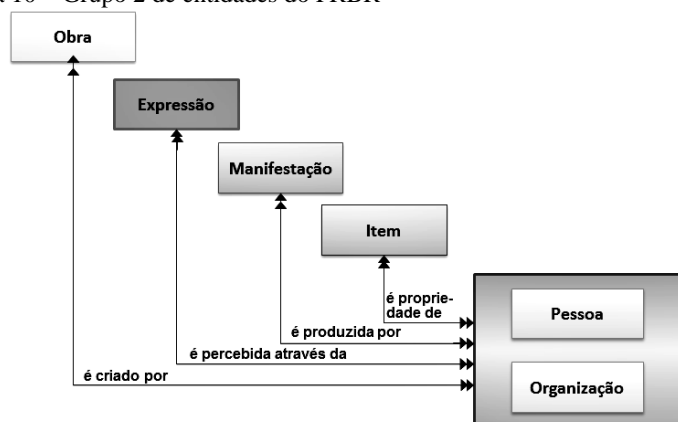
Figura 9 – Grupo 1 de entidades do FRBR



Fonte: Adaptado de IFLA (2009).

As entidades do grupo 2 representam os responsáveis pelo conteúdo intelectual ou artístico, a sua produção física e a disseminação ou que tem a custódia das entidades do grupo 1. As entidades que integram esse grupo são, conforme Figura 10: **Pessoa** é um indivíduo; e **Organização** é uma organização, um grupo de indivíduos e/ou organizações.

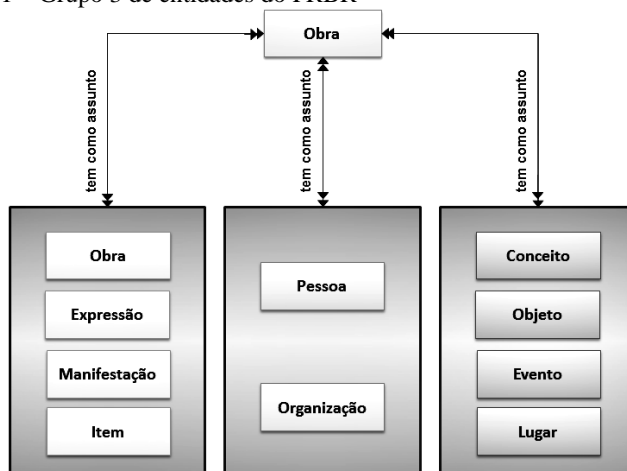
Figura 10 – Grupo 2 de entidades do FRBR



Fonte: Adaptado de IFLA (2009).

As entidades do grupo 3 representam as relações de assuntos entre a obra e as entidades do grupo 1 e do grupo 2, e tem o acréscimo de quatro entidades – Figura 11: **Conceito** é uma noção resumida ou ideia; **Objeto** é a coisa material; **Evento** é uma ação ou ocorrência; e **Lugar** é uma localização.

Figura 11 – Grupo 3 de entidades do FRBR



Fonte: Adaptado de IFLA (2009).

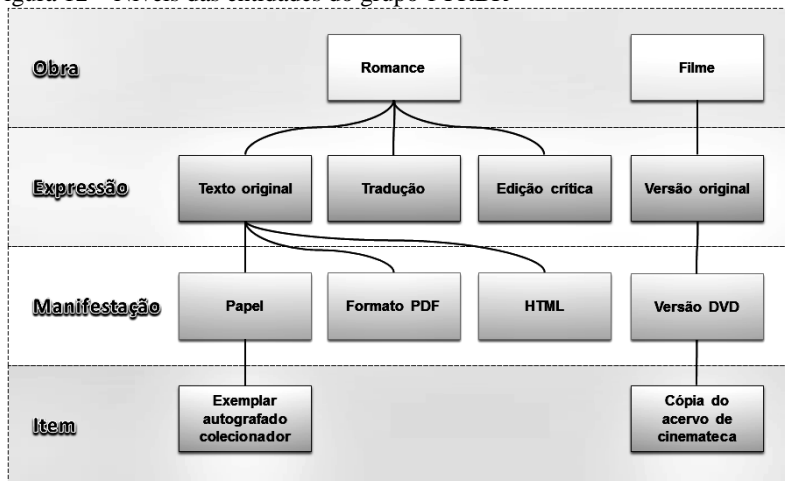
Os relacionamentos da entidade *obra* com os grupos 1 e 2 explicita que uma *obra* pode ter como assunto uma outra obra, uma expressão de uma obra, uma manifestação de uma obra ou, ainda, sobre um item específico, assim como pessoa ou organização.

Para este trabalho, a desambiguação que o grupo 1 do modelo FRBR permite tem particular interesse. Com a definição dos conceitos das entidades do grupo 1, as FRBR permitem diferenciar, por exemplo, as diferentes acepções da palavra “livro” (LE BOEUF, 2007; TILLET, 2003):

- quando se diz “livro” para fazer referência a um objeto físico, como um exemplar comprado em uma livraria ou emprestado de uma biblioteca, os FRBR chamam esse objeto de *Item*;
- quando se diz “livro” no sentido da “publicação”, como um conjunto de objetos físicos, como na frase “quero ir à livraria comprar o livro x”, o modelo FRBR emprega o termo *Manifestação*;
- no contexto de “quem traduziu o livro”, “esta é a versão do livro x”, são as *Expressões* de uma *Obra*, como texto original, texto traduzido, edição ilustrada;
- no mais alto nível de abstração, no sentido da criação, por exemplo, “quem escreveu o livro”, fazendo referência às diferentes versões de uma criação,

consideradas como a mesma, o modelo FRBR refere-se a esse conceito de *Obra*.

Figura 12 – Níveis das entidades do grupo 1 FRBR



Fonte: Adaptado de Tillett (2004).

A Figura 12 ilustra o exemplo de Tillett (2004), no qual a *obra* “E o vento levou” de Margareth Mitchel, que teve várias *expressões*, incluindo o texto original, tradução, edição crítica, e filme. Foi *manifestada* como livro em papel e duas edições eletrônicas, uma em PDF e outra em HTML. Em *item*, estão as cópias específicas, a exemplo de um exemplar de livro autografado de um colecionador ou uma cópia do filme no acervo de uma cinemateca.

Ao estabelecer os FRBR, a intenção do grupo de estudos da IFLA é reduzir a ambiguidade e permitir a interoperabilidade entre os sistemas de catalogação bibliográfica à medida que forem “FRBR-rizados”.

Para melhor compreensão, é necessária a contínua reflexão do conhecimento como um fenômeno organizacional, e distingui-lo de dados e informações (FAHEY; PRUSAK, 1998).

2.1.3 Dado – Informação – Conhecimento

Em se tratando do assunto conhecimento, um recorrente debate é a polêmica em torno do entendimento, diferenças e semelhanças entre dado, informação e conhecimento, especialmente entre os dois últimos que, frequentemente, são utilizados como sinônimos.

Fahey e Prusak (1998) consideram que o primeiro dos onze erros fatais da GC é a falta de clareza e definição de conhecimento. Afirmam que, se o conhecimento não for algo diferente de dado ou informação, então, não haveria nada de novo ou interessante na GC e que a relutância não está somente na distinção entre o conhecimento, dados e informações, mas também em considerar quais as implicações dessas distinções – o que é mais importante.

A reflexão sobre conceitos e as diferenças entre eles é a essência no processo de aprendizagem e ignorá-lo é um erro crítico que contribui diretamente para os outros erros (FAHEY; PRUSAK, 1998).

A equivalência entre dado, informação e conhecimento propicia coisificar o conhecimento, o que pode induzir administradores a pensarem que a Gestão do Conhecimento na organização está encaminhada ao adquirir ferramentas de gestão, e manipulação de dados e informações.

Em muitas empresas, o conhecimento simplesmente tornou-se outro objeto a ser gerenciado (FAHEY; PRUSAK, 1998). Numa “FRBR-rização” do assunto, seria como dar ênfase à gestão do *item* e não da *obra*.

A hierarquização do dado-informação-conhecimento-sabedoria, conhecido como hierarquia DIKW (*data-information-knowledge-wisdom*), há tempo, é objeto de estudos, conceitualizações e discussões sem, no entanto, ter convergido para um consenso.

As várias vertentes ficam visíveis na denominação dessa visão hierárquica. Pode ser denominada hierarquia do conhecimento, hierarquia da informação, pirâmide do conhecimento, pirâmide informacional, pirâmide DIKW (FRICKÉ, 2009; ROWLEY, 2007; SHARMA, 2008; URDANETA, 1992), entre outras variações (Figura 13).

Figura 13 – Pirâmide DIKW



Fonte: Elaborado pelo autor.

Segundo Rowley (2007), a pirâmide DIKW é um fundamental e amplamente reconhecido modelo na literatura acerca da informação e conhecimento, frequentemente, a origem é atribuída ao artigo de Ackoff de 1989, denominado *From data to wisdom*.

A hierarquia DIKW supõe, implicitamente, que dados podem ser usados para criar informação, informação pode ser usada para criar conhecimento e conhecimento pode ser usado para criar sabedoria (ROWLEY, 2007).

Na definição de Ackoff (apud ROWLEY, 2007):

- **Dados** são símbolos que representam propriedades de objetos, eventos e seus ambientes são produtos da observação;
- **Informação** está contida em descrições, respostas e perguntas que iniciam com palavras como quem, o que, quando, e quantos. Informações são inferidas de dados;
- **Conhecimento** é *know-how*, é o que torna possível a transformação da informação em instruções. Pode ser obtida pela transmissão daquele que a possui ou se extrai da experiência; e

Sabedoria é a habilidade para aumentar a efetividade. Adiciona valor que requer julgamento.

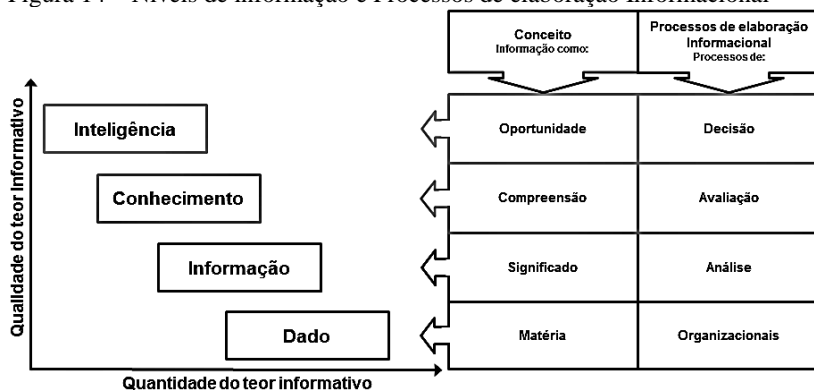
Com ênfase na informação, Urdaneta (1992, pp. 99-100) argumenta que o ordenamento dos níveis da pirâmide (que denomina de pirâmide informacional) não é arbitrário. A hierarquização tem como base as noções *qualidade versus quantidade* na qual o atributo de inteligência tem mais o caráter qualitativo, no sentido de que depende mais da qualidade da informação do que sua quantidade. No caso dos

dados, a quantidade de informação seria mais decisiva do que a sua qualidade (URDANETA, 1992).

Urdaneta (1992) chama o topo da pirâmide de inteligência e define cada camada como:

- **Dado** é “informação como matéria”, no sentido de que são registros icônicos, simbólicos, sîgnico por meio dos quais são representados: coisas, fatos, conceitos ou instruções;
- **Informação** é “informação como significado”, no sentido de que são dados ou materiais informacionais relacionados ou estruturados de forma potencialmente significativa;
- **Conhecimento** é “informação como compreensão”, no sentido de que são estruturas informacionais que, ao internalizar-se, se integram a sistemas de relacionamento simbólico de mais alto nível e permanência. Trata-se, portanto, de informação que pode ser entendida por aquele que tem o domínio da linguagem utilizada, mas que somente pode ser compreendido por aquele que conta com uma base de conhecimento para tal; e
- **Inteligência** é “informação como oportunidade” no sentido de que são estruturas de conhecimento que, sendo contextualmente relevantes, permitem a intervenção da realidade.

Figura 14 – Níveis de informação e Processos de elaboração Informacional



Fonte: Baseado em Urdaneta (1992).

Urdaneta (1992, pp. 105-106), em sua proposta, apresenta um modelo de agregação de valor que ordena os quatro níveis de informação em função de quatro classes de processos de elaboração informacional (Figura 14):

- os processos organizacionais implicam: agrupamento, classificação, relacionamento, formatação, sinalização e apresentação dos dados para convertê-los em informação;
- os processos de análise implicam: discriminação, qualificação, validação, comparação, interpretação e sintetização da informação para convertê-lo em conhecimento;
- os processos de avaliação implicam: apresentação de opções, vantagens e desvantagens do conhecimento, que pode ser convertido em inteligência; e
- os processos decisórios implicam: alinhamento de metas, comprometimento, negociação ou escolha de inteligência que possibilite a ação corretiva do problema.

A frequente confusão sobre os significados e as diferenças entre dado, informação e conhecimento tem como consequência dispêndios em iniciativas tecnológicas, com raros resultados correspondentes.

É importante entender que os conceitos de dado, informação e conhecimento não são intercambiáveis, e que a falha e o sucesso organizacional podem depender da clareza que a organização possui em relação ao (DAVENPORT; PRUSAK, 2000):

- que é necessário;
- que possui; e
- que pode e não pode fazer com eles.
- Na hierarquia dado-informação-conhecimento de Davenport e Prusak (2000), cada elemento é caracterizado conforme o Quadro 11.

Quadro 11 – Dado, Informação e Conhecimento

Dado	Informação	Conhecimento
Conjunto de fatos discretos e objetivos de eventos.	Dado contextualizado, calculado, categorizado, corrigido, condensado.	Composto fluido de experiência, valores, informação contextual e percepções especializadas.
Frequentemente estruturado e	Qualitativamente avaliada pela qualidade e utilidade	Embutido com frequência em documentos,

armazenado em sistemas tecnológicos.	da informação.	repositórios, rotinas, processos, práticas e normas.
Gerenciado quantitativamente em termos de custo, velocidade e capacidade.	Quantitativamente tende a ser medida pela conectividade e transações.	Difícil de capturar ou entender completamente em termos lógicos.
Sem significado inerente.	Dotado de significado – relevância e propósito.	Informação transforma em conhecimento por meio da comparação, análise de consequências, conexão e conversação.
Descreve somente parte do que aconteceu	Mensagem com remetente e destinatário.	A criação de conhecimento ocorre nas pessoas e entre pessoas.
Material bruto necessário para a criação da informação.	Tenciona mudar a forma de o destinatário perceber algo, ter um impacto no seu julgamento e comportamento.	É transferido por meio de mídia estruturada, como livros e documentos, e de pessoa a pessoa.

Fonte: Baseado em Davenport e Prusak (2000).

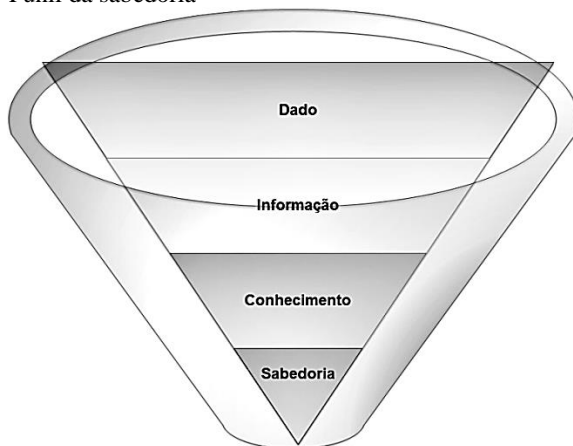
Numa análise de definições adotadas em artigos, Rowley (2007) conclui que há consenso de conceitos de dado, informação e conhecimento, reafirmando o conceito de uma hierarquia que liga os conceitos dos três elementos. Observa, também, que há menor concordância na natureza do processo que converte dados em informação e informação em conhecimento, mas a distinção entre informação e conhecimento é confusa. A autora argumenta que, se conhecimento é uma propriedade de pessoas e incorpora um entendimento prévio, experiência e aprendizagem, fica difícil argumentar que o conhecimento explícito gravado em documentos e sistemas de informações é mais ou menos que informação.

Rowley (2007) observa ainda em sua análise que a hierarquia DIKW sugere que:

- há mais dado do que informação, que conhecimento, que sabedoria. Seria isto um desejo aceitável ou simplesmente a expressão do estado atual?
- a hierarquia, com sua larga base de dados, é segura e estável; e

- sabedoria é somente alcançada após muito processamento de dados, informação e conhecimento, e o processo tem início com dados.
- A autora conclui, na sua análise, que quiçá a hierarquia DIKW seria mais evocativa se invertida, como na Figura 15.

Figura 15 – Funil da sabedoria



Fonte: Adaptado de Rowley (2007).

No funil da sabedoria, os dados naturalmente seriam mais concentrados e o edifício todo é delicadamente balanceado sobre a sabedoria e desmoronará se não houver sabedoria suficiente (ROWLEY, 2007).

O modelo DIKW também é alvo dos críticos, entre eles, Frické (2009) aponta que a hierarquia DIKW tem ascendência no positivismo ou operacionalismo, realiza generalizações estatística, não permite inferências e encoraja metodologias sem inspiração. A visão que dados são coletados, promovidos à informação e informação responde questões, estimula a coleção de dados sem sentido na esperança de um dia ser elevada à informação (FRICKÉ, 2009).

Embora critique o modelo DIKW, Frické deixa claro que não defende o seu abandono, mas que deve deixar de ser o cânone da ciência da informação, opinião compartilhada por outro crítico do modelo Bernstein (2009) que propõe uma reavaliação da hierarquia e não o banimento.

Bernstein (2009) revisita a hierarquia DIKW, propõe uma antítese ao modelo e apresenta a contraparte negativa dos termos da hierarquia (Quadro 12).

Quadro 12 – Contraparte negativa do DIKW

DIKW	Contraparte negativa
<i>Dado</i>	<i>Ausência ou carência de dados; dados inexistentes.</i>
Informação	Informação errada; desinformação; erro
Conhecimento	Ignorância
Sabedoria	Insensatez; estupidez

Fonte: Baseado em Bernstein (2009).

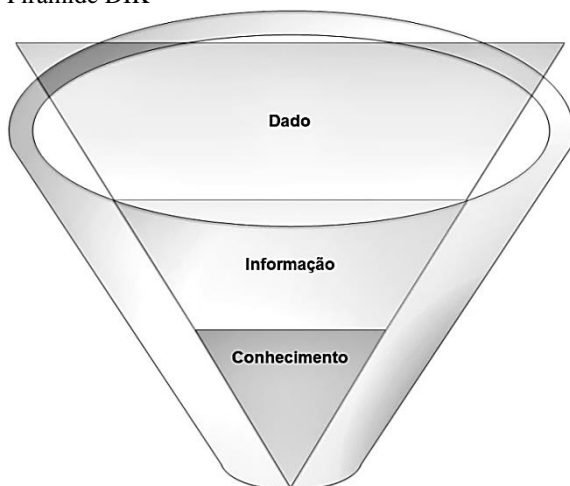
Bernstein (2009) argumenta que o oposto do dado é uma possibilidade lógica, tendo em vista que um dado pode não ser dado para tudo, isto é, pode ser verdadeiro num contexto e falso em outro. A contraparte negativa da informação pode ser diversa. Um erro é um lapso feito inadvertidamente ou por ignorância, enquanto uma informação errada (*misinformation*) é incorreta ou um falso registro, e a desinformação é a disseminação deliberada de informação falsa com a intenção de influenciar a opinião (BERNSTEIN, 2009).

Nesse estudo da antítese da hierarquia DIKW, Bernstein (2009) argumenta que, embora possa ser identificada a contraparte negativa de cada elemento da pirâmide, esses elementos opostos não formam uma pirâmide, pois o oposto de dado pode não conduzir ao oposto de informação, etc. O autor enfatiza que a proposta é o desenvolvimento de um conceito compreensivo, conceito de “*nonknowledge*” – a antítese do conhecimento.

O *nonknowledge* é parte do mundo e apresenta um desafio na organização do conhecimento, particularmente na diferenciação do *nonknowledge* como um tópico e a investigação na área de estupidez, erros, informação errada, falta de dados, encontra importantes conexões entre estas áreas, apesar de não ser um perfeito espelho do DIKW (BERNSTEIN, 2009).

Para o presente trabalho estabelecer clareza nos conceitos e nas relações informação-conhecimento, é essencial, e como o foco é o conhecimento, o funil poderá ser limitado em três níveis com a organização apoiada sobre o conhecimento devidamente organizado e selecionado.

Figura 16 – Pirâmide DIK



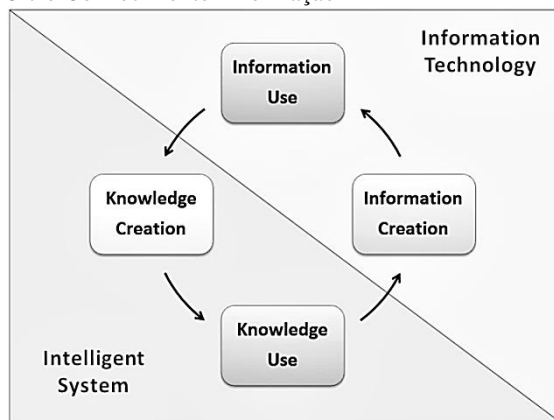
Fonte: Elaborado pelo autor baseado em Rowley (2007).

No modelo de Blumentritt e Johnston (1999), das quatro categorias propostas (Quadro 9), o conhecimento codificado é equivalente à informação, e afirmam que conhecimento e informação são diferentes, pois conhecimento requer o contexto de um sistema inteligente, assunto já abordado e ilustrado na Figura 4, mas que merece ser retomado por ser um dos aspectos centrais deste trabalho.

Na proposta dos autores, a fronteira entre informação e conhecimento pode ser claramente estabelecida, pois o conhecimento somente existiria dentro de um sistema inteligente, e a informação é que pode ser capturada e transmitida na forma digital, portanto, em toda comunicação com o envolvimento de um sistema inteligente, o que é comunicado não é o conhecimento, mas o resultado de um processo de conhecimento, e seria somente uma parte do que foi envolvido no processo.

Portanto, na circularidade da informação e conhecimento, a transmissão, difusão ou compartilhamento do conhecimento irá sempre requerer que ele seja traduzido para informação (Figura 17).

Figura 17 – Ciclo Conhecimento-Informação



Fonte: Adaptado de Blumentritt e Jonhnston (1999).

Para a informação voltar ao estado do conhecimento, ela deve fazer sentido no contexto do usuário, construído da experiência, da cultura, das relações sociais e da educação, permitindo que ele crie ligações entre diferentes conjuntos de informação (BLUMENTRITT; JOHNSTON, 1999).

Acerca da relação informação-conhecimento, Blumentritt e Johnston (1999) concluem:

- o conhecimento requer o contexto de um meio inteligente;
- o conhecimento resulta da combinação de informações para formar um quadro mental que serve como base;
- a informação é resultado de um processo de conhecimento;
- a informação dá suporte ao conhecimento, mas não o substitui;
- um sistema de Gestão da Informação não possibilita gerir conhecimento – para gerenciar conhecimento, ferramentas devem contemplar as características especiais do conhecimento, incluindo a complexidade da estrutura do conhecimento;
- a Gestão da Informação já está bem desenvolvida, a Gestão do Conhecimento ainda está na sua infância.

Da mesma forma que Bernstein (2009) considera que o dado e o “não dado” uma possibilidade lógica, analogamente, a relação

informação-conhecimento é uma possibilidade lógica. O conceito relacionado a cada termo depende do contexto e da vertente do pensamento, variando de sinônimos a coisas distintas.

No modelo DIKW de Urdaneta (1992), cada camada da hierarquia é uma forma de manifestação da informação, de forma contrastante, as categorias propostas por Blackler, Lam, Musgrave, Lundvall, Collins, Millar e Fleck fazem somente menção ao conhecimento.

Blumentritt e Johnston denomina todas as categorias como conhecimento, mas assume que o conhecimento codificado é equivalente à informação.

Observa-se, nas diferentes abordagens, que o foco determina a densidade de cada assunto. O foco na informação tende a transformá-la em uma entidade complexa, com diferentes atributos e categorias, e, nesse caso, o conhecimento fica restrito ao conteúdo incorporado na mente do conhecedor.

2.2 GESTÃO DO CONHECIMENTO

Aceitar o conhecimento como a chave para a vantagem competitiva causa muitos desafios para o mundo corporativo e seus gestores, que, acostumados a agrupar recursos tangíveis, físicos, financeiros e humanos, e competir sobre preço, qualidade e serviços a clientes, têm dúvidas de como gerir algo escorregadio e fugidio como o conhecimento (BLUMENTRITT; JOHNSTON, 1999).

As rápidas e constantes transformações, isto é, a mudança como um estado contínuo pode ter consequência funesta para as organizações que não tiverem um processo de atualização contínua do conhecimento. O conhecimento organizacional deteriora, torna-se obsoleto, podendo levar a decisões errôneas (FAHEY; PRUSAK, 1998).

Nesse contexto, uma nova disciplina ocupa a atenção de acadêmicos e profissionais – a Gestão do Conhecimento (GC) – que possui várias vertentes e interpretações, e vários modelos. O próprio entendimento do que se trata a GC evidencia essa diversidade.

Para Gupta, Iyer e Aronson (2000), GC é o processo que ajuda as organizações na busca, seleção, organização, disseminação e transferência de informações importantes e competências necessárias na resolução de problemas, na aprendizagem, no planejamento estratégico e na tomada de decisão.

Na visão de Davies (2000), a GC tem como meta explorar a informação organizacional para grande produtividade, novo valor e

competitividade crescente e, também, para encorajar a permuta do conhecimento tácito.

Para Wiig (1997), o propósito geral da GC é maximizar a efetividade e o retorno dos ativos de conhecimento e renová-los constantemente. A GC tem como objetivo executar as ações da empresa tão inteligentemente possível para assegurar a viabilidade e o sucesso geral, isto é, realizar o melhor valor de seus ativos de conhecimento (WIIG, 1997).

Na visão de Nonaka e Takeuchi (1997), o foco da GC é a criação de conhecimento. O conhecimento individual é traduzido em conhecimento organizacional por meio do fluxo do conhecimento tácito para explícito. Essa tradução ocorre por meio de quatro modos de conversão do conhecimento, que são: socialização, externalização, combinação e internalização, pode ser observada na Figura 18. A espiral do conhecimento é o núcleo da teoria de Nonaka e Takeuchi.

Figura 18 – Espiral do conhecimento



Fonte: Adaptado de Nonaka e Takeuchi (1997, p.80).

Como citado na seção 2.1, tanto Nonaka e Takeuchi (1997, p.62) quanto Lam (1998), adotam as dimensões ontológica e epistemológica para o conhecimento. Embora o estabelecimento das duas dimensões da criação do conhecimento possa, em primeira análise, ter a conotação puramente filosófica e acadêmica, a assunção dessa visão pode levar a diferentes resultados.

A visão de mundo cartesiana do sujeito e objeto, do conhecedor e conhecido, reduziu a organização a um mecanismo de processamento de informações recebidas do ambiente externo na tentativa de adaptação às novas circunstâncias, não explicando a inovação, que não se limita ao processamento das informações de fora para dentro, mas na criação de novos conhecimentos e informações, no fluxo inverso de dentro para fora, recriando seu meio (NONAKA; TAKEUCHI, 1997, p. 61).

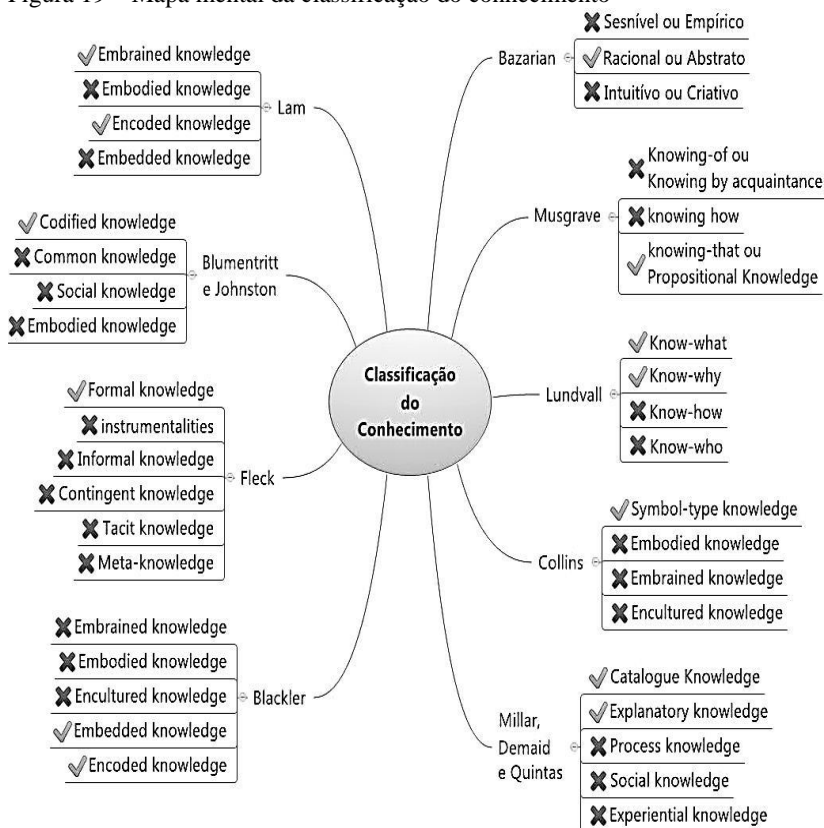
As teorias organizacionais reconhecem, atualmente, três áreas distintas nas quais a criação e o uso da informação têm papel estratégico na determinação da capacidade de crescimento e adaptação da organização: entender as mudanças em seu ambiente; e criar novos conhecimentos para inovação e na tomada de decisão (CHOO, 1996).

O conhecimento e a GC são conceitos complexos e multifacetados (ALAVI; LEIDNER, 2001), e as várias visões e modelos de GC são consequências do conceito adotado para o seu objeto central – o conhecimento já explorado no item 2.1.

2.2.1 Conhecimento e GC: Síntese dos aspectos de interesse

A transitoriedade do ser humano e o consequente caráter impermanente do conhecimento nele contido levaram os pensadores, pesquisadores e profissionais ao desafio de como o conhecimento pode ser transferido. Essa preocupação de identificar o conhecimento que pode ser transferido e o de difícil transferência fica visível nas classificações propostas pelos autores consultados. Sintetizados no mapa mental (Figura 19), os conhecimentos considerados mais fáceis de serem transferidos foram identificados com o sinal ✓ e os conhecimentos considerados impossíveis ou difíceis de serem transferidos com o sinal ✕.

Figura 19 – Mapa mental da classificação do conhecimento



Fonte: Elaborado pelo autor.

O objeto de pesquisa do presente trabalho é especificamente o conhecimento codificado – no conceito do *Encoded Knowledge* de Blackler (1995) e Lam (1998) e do *codified knowledge* de Blumentritt e Johnston (1999) – e registrado em suporte digital como forma de transferência do conhecimento.

A visão da Teoria do Conhecimento objetivo de Popper permite a clara separação entre o objeto físico (mundo 1) e o conhecimento nele registrado (mundo 3). No caso dos objetos digitais, essa distinção é importante, considerando que o suporte físico (ótico ou magnético) não é o mais importante, desde que seja assegurado que o objeto digital esteja preservado em um suporte. A Teoria do Conhecimento objetivo

de Popper admite a existência do conhecimento sem o conhecedor, sem o qual não seria possível preservar o conhecimento.

O modelo FRBR oferece uma desambiguação que é de grande importância, especialmente com os conceitos desenvolvidos pelas entidades do grupo 1: *Obra, Expressão, Manifestação e Item*. A “FRBR-rização” permite a clara distinção entre o objeto físico e a criação intelectual, isto é, o conhecimento.

A questão central deste trabalho é a preservação do conhecimento codificado, para isso, é necessário assegurar a integridade do suporte em que o conhecimento está armazenado. Isso não significa que o suporte original deva ser preservado, pelo contrário, ele deve ser continuamente renovado, para novas mídias com tecnologia atualizada.

A preservação do conhecimento codificado armazenado em objetos digitais não pode ser confundida com a preservação física da mídia. É uma das diferenças da preservação do conhecimento registrado em papel, no qual a preservação do conhecimento se confunde com a preservação do objeto físico.

Esse discernimento é importante para a elaboração e o entendimento do modelo conceitual da solução.

As práticas e os pensamentos relacionados à Gestão do Conhecimento têm influência de várias disciplinas, destacando-se: a filosofia (definindo o conhecimento); a ciência cognitiva (para entender os trabalhadores do conhecimento); as ciências sociais (entendendo a motivação, pessoas, interações, cultura e ambiente); a ciência da informação (construindo as capacidades relacionadas ao conhecimento); a Engenharia do Conhecimento (na elicitação e codificação do conhecimento); a ciência administrativa (otimizando operações e integrando-o à empresa) (KAKABADSE N.; KAKABADSE A.; KOUZMIN, 2003).

A literatura revela muitos modelos de gestão do conhecimento: filosófico, cognitivo, rede, comunidade, quântico. O modelo cognitivo tem recebido considerável atenção (KAKABADSE N.; KAKABADSE A.; KOUZMIN, 2003).

Para o modelo cognitivo, o conhecimento é um ativo, objetivamente definido e codificado como conceitos e fatos, tem foco na captura e no armazenamento do conhecimento, necessita ser gerido e tem a tecnologia como uma camada crítica do modelo (KAKABADSE N.; KAKABADSE A.; KOUZMIN, 2003).

O modelo cognitivo de Gestão do Conhecimento está, portanto, alinhado com a necessária visão para o desenvolvimento deste trabalho – a gestão e preservação do conhecimento codificado.

2.3 ENGENHARIA DO CONHECIMENTO

A Engenharia do Conhecimento fornece métodos para compreender as estruturas e os processos utilizados pelos trabalhadores do conhecimento, permitindo a integração da tecnologia da informação no apoio à Gestão do Conhecimento corporativo, identificando as oportunidades e os gargalos na forma como elas aplicam os seus recursos de conhecimento (SCHREIBER *et al.*, 2000).

A Engenharia do Conhecimento, geralmente, envolve cinco passos na transferência do conhecimento humano em um sistema baseado em conhecimento (KENDAL; CREEN, 2007):

- 1 aquisição do conhecimento, que envolve a obtenção do conhecimento de várias fontes como especialistas, livros, vídeos, bases de dados e Internet;
- 2 validação do conhecimento, que é a verificação com testes de qualidade adequada;
- 3 representação do conhecimento, que envolve a produção de um mapa de conhecimento e a codificação numa base de conhecimento;
- 4 inferência, que é formar ligações do conhecimento com o *software* para que o sistema baseado em conhecimento possa ajudar na decisão do usuário; e
- 5 explanação e justificação, que envolve projetos de *softwares* adicionais, principalmente para ajudar o computador a responder questões postadas pelo usuário e, também, para mostrar como uma conclusão foi alcançada com o uso da base de conhecimento.

A Engenharia do Conhecimento define um processo no qual especialistas de domínio e engenheiros do conhecimento constroem uma base de conhecimento, que é representada em uma linguagem de representação de conhecimento e ferramentas adequadas (AUER, 2009). Ela condensa a informação elicitada em conhecimento, por meio da estruturação, formalização e operacionalização (AUER, 2009).

2.3.1 Ontologia

Na Engenharia do Conhecimento, a ontologia é definida como o objeto de consenso de humanos e um objeto formal que permite o uso por agentes de *softwares* (AMARDEILH, 2009).

Ontologia é um termo que tem origem na Filosofia e refere-se ao estudo sistemático da existência (GÓMEZ-PÉREZ, 1999).

A definição frequentemente citada de ontologia é de Gruber (1995; p. 1): “Uma ontologia é uma especificação explícita de uma conceitualização”.

Na computação, uma ontologia é uma descrição formal e explícita de conceitos de um domínio do discurso (NOY; MCGUINNESS, 2001) que define um conjunto de primitivas representacionais, que são as classes (os conceitos), as propriedades e os relacionamentos entre as classes (GRUBER, 2009).

Os indivíduos são as instâncias de classes que representam os objetos no domínio (HORRIDGE *et al.*, 2007). Uma base de conhecimento é a ontologia com um conjunto de indivíduos (NOY; MCGUINNESS, 2001).

A ontologia é um artefato projetado para permitir a modelagem de conhecimento em algum domínio e permite especificar um vocabulário para fazer asserções que podem ser utilizadas por *software*.

O Protégé é uma plataforma de código aberto, desenvolvida no *Stanford Center for Biomedical Informatics Research* da *Stanford University School of Medicine*. É um conjunto de ferramentas para a construção de modelos de domínio e aplicações baseada em conhecimento com ontologias, com o suporte à criação, visualização e manipulação de ontologias⁷.

Existem várias linguagens para a criação de ontologia. Para este estudo, a opção foi pela *Web Ontology Language* (OWL) do *World Wide Web Consortium* (W3C).

A versão Protégé-OWL permite a construção de ontologias na linguagem OWL – *Web Ontology Language*, uma recomendação do consórcio W3C⁸.

Ontologias OWL podem ser categorizadas em três sublinguagens: OWL-Lite, a menos expressiva; a OWL Full, a mais expressiva e a OWL-DL, com expressividade entre a Lite e a Full (HORRIDGE *et al.*, 2007).

2.3.2 Mapas Conceituais

Mapas conceituais são ferramentas gráficas para organizar e representar o conhecimento (NOVAK; CAÑAS, 2008). São diagramas

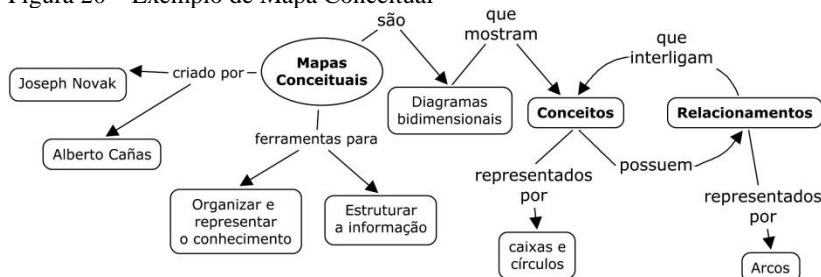
⁷ <http://protege.stanford.edu/overview/>

⁸ <http://www.w3.org/TR/owl-overview/>

bidimensionais que mostram as relações entre conceitos que derivam da própria estrutura conceitual do domínio mapeado (MOREIRA, 2006, p. 46). É uma maneira alternativa de estruturar a informação (AMORETTI; TAROUÇO, 2000).

O mapa conceitual foi desenvolvido no projeto de pesquisa de Novak, na Universidade de Cornell, no início da segunda metade do século XX, a partir da necessidade de encontrar uma melhor maneira de representar compreensões conceituais das crianças, e de observar as mudanças na estrutura conceitual e proposicional que constroem aqueles entendimentos (NOVAK; CAÑAS, 2006). A Figura 20 apresenta um exemplo de mapa conceitual com conceitos e suas relações.

Figura 20 – Exemplo de Mapa Conceitual



Fonte: Yamaoka e Gauthier (2013).

Os mapas conceituais exibem, graficamente, o conhecimento que é composto por conceitos (representados por caixas ou círculos). Os conceitos são interligados por arcos e rotulados com palavras que formam proposições ou frases entre pares de conceitos (CAÑAS *et al.*, 2005).

No início, os mapas conceituais eram construídos manualmente, mas as facilidades para elaborar e compartilhar foram imensamente ampliadas com o desenvolvimento do *software* CmapTools pelo Institute for Human and Machine Cognition (IHMC)⁹.

O IHMC não só dissemina o CmapTools, mas, também, a teoria, o fundamento basilar e as orientações de como construir bons mapas conceituais. Os passos recomendados pelo IHMC (2009) para construir bons mapas conceituais são:

⁹ <http://cmap.ihmc.us/>

- 1 Definir o contexto para o mapa conceitual, estabelecendo a questão focal que será usada para identificar os conceitos relevantes;
- 2 Fazer uma lista de 15 a 25 conceitos;
- 3 Ordenar os conceitos;
- 4 Iniciar a construção do mapa com 1 a 4 conceitos mais gerais;
- 5 Escolher as palavras para relacionar os conceitos procurando formar boas proposições;
- 6 Continuar a construção da hierarquia de conceitos;
- 7 Buscar por possíveis ligações cruzadas que mostram relacionamentos entre conceitos em duas seções distintas do mapa;
- 8 Reposicionar e refinar a estrutura do mapa.

2.4 ARQUIVOLOGIA

Na literatura, não há consenso sobre a própria denominação da Arquivologia – o termo arquivística é também utilizado como sinônimo. Igualmente sem consenso, a Arquivologia ora é designada como ciência (*Archival Science*) ora como disciplina (disciplina arquivística). Para este trabalho, foi adotado o termo Arquivologia em consonância com o Dicionário Brasileiro de Terminologia Arquivística (ARQUIVO NACIONAL, 2005), no qual esse é indicado como termo preferencial.

A Arquivologia emergiu da necessidade e do saber prático (*savoir-faire*) acumulado ao longo dos anos (ROUSSEAU; COUTURE, 1998), que “estuda as funções do arquivo¹⁰ e os princípios e técnicas a serem observados na produção, organização, guarda, preservação e utilização dos arquivos¹¹” (ARQUIVO NACIONAL, 2005).

A ciência arquivística tem dois conceitos principais: arquivo e documento arquivístico (THOMASSEN, 2001).

O termo arquivo vem de *archion*, utilizado pelos gregos no século III ou II a.C. para indicar “*governmet palace, general administrator, office of the magistrate, records office, original records, repositor for original records, authority*” (ROUSSEAU; COUTURE, 1998).

O termo arquivo é polissêmico, podendo denotar:

¹⁰ Arquivo como instituição

¹¹ Arquivo como conjunto de documentos

- 1 Conjunto de documentos produzidos e acumulados por uma entidade coletiva entidade coletiva entidade coletiva entidade coletiva, pública ou privada, pessoa ou família, no desempenho de suas atividades, independentemente da natureza do suporte.
- 2 Instituição ou serviço que tem por finalidade a custódia, o processamento técnico, a conservação e o acesso a documentos.
- 3 Instalações nas quais funcionam arquivos.
- 4 Móvel destinado à guarda de documentos.

(ARQUIVO NACIONAL, 2005)

O uso de termos diferentes permite distinguir a instituição e o material de que se ocupa (SCHELLENBERG, 2006). Na obra “Arquivos Modernos”, o autor adota a palavra “arquivo” para designar a instituição e a expressão “material de arquivo” ou “arquivos” para designar o material.

A antiga civilização grega é, provavelmente, na qual se originaram os arquivos como instituição (SCHELLENBERG, 2006).

Os arquivos públicos foram instituídos por quatro motivos: o primeiro foi a necessidade de incrementar a eficiência governamental; o segundo, de ordem cultural, os arquivos públicos constituem um tipo de fonte de cultura; o terceiro foi de interesse pessoal, visando determinar relações sociais, econômicas e políticas; e o quarto motivo foi de ordem oficial, pois documentos são necessários às atividades do governo, refletindo sua origem e seu crescimento, e constitui a principal fonte de informação das suas atividades, de comprovação de obrigações legais e financeiras (SCHELLENBERG, 2006). Portanto, os arquivos públicos atuam na política de arquivos, no recolhimento, na preservação e na divulgação do patrimônio documental sob sua custódia.

2.4.1 O repensar da ciência arquivística

A Arquivologia tradicional tem base na modelo de ciclo de vida (*life-cycle model*) que possui defensores na academia e adotado por instituições arquivísticas importantes, como a dos Estados Unidos, Reino Unido, França, Itália e também do Brasil.

O modelo de ciclo de vida surgiu no final dos anos 30 no Arquivo Nacional dos Estados Unidos, desenvolvidos, inicialmente, com Phillip Coolidge Brooks e Emmet Leahy (CHELL, 2007). No pioneiro artigo sobre seleção para a preservação de documentos arquivísticos, Brooks (1940) apresenta narrativamente um diagrama do ciclo de vida dos

documentos, da criação e recepção, seleção, descarte ou preservação permanente.

O modelo de ciclo de vida é baseado na premissa que é possível dividir a vida de um documento arquivístico em oito estágios distintos, quatro na primeira fase – criação ou recepção da informação na forma de documento, classificação, manutenção/uso e transferência ou destruição – e quatro na segunda fase – seleção/aquisição dos documentos de arquivo por uma instituição, descrição, preservação e referência, e uso por pesquisadores (ATHERTON, 1985).

Nascida a partir de diferentes culturas, motivações e necessidades em diferentes localização e tempo, quiçá por isso a Arquivologia não tem consenso mesmo no que se refere aos princípios que a rege. Em alguns momentos, o domínio do empirismo e o senso comum parecem ser a marca da área (SOUSA, 2008).

Segundo Rousseau e Couture (1998), três princípios constituem o fundamento da arquivística: o princípio da territorialidade; o princípio da proveniência; e a abordagem das três idades.

O princípio da territorialidade é o “Princípio derivado do princípio da proveniência e que defende que os arquivos devem ser conservados nos serviços de arquivo do território no qual foram produzidos ou, melhor ainda, na instituição produtora do fundo” (ROUSSEAU; COUTURE, 1998, p. 295).

O princípio da proveniência é basilar na Arquivologia tradicional:

O princípio da proveniência que comporta duas vertentes define-se como o princípio fundamental segundo o qual os arquivos de uma mesma proveniência não devem ser misturados com os de outra proveniência e devem ser conservados segundo a sua ordem primitiva, caso exista ou o princípio segundo o qual cada documento deve ser colocado no fundo donde provém e, nesse fundo, no seu lugar de origem.

[...] O princípio da proveniência e o seu resultado, o fundo de arquivo, impõe-se à arquivística, uma vez que esta tem por objetivo gerir o conjunto das informações geradas por um organismo ou por uma pessoa no âmbito das atividades ligadas à missão, ao mandato e ao funcionamento do dito organismo ou ao funcionamento e à vida da referida pessoa (ROUSSEAU; COUTURE, 1998, p. 80-82).

O fundo de arquivo é um contexto particular, um agrupamento intelectual de informações registradas, em que os arquivos devem se desenvolver, permitindo, assim, que o contexto de realização de um acontecimento seja recriado, seja a vida de uma personagem ou de um organismo, situando o conjunto de documentos nesse contexto, no tempo e no espaço (ROUSSEAU; COUTURE, 1998).

A abordagem das três idades estabelece as etapas de vida dos documentos: idade ativa, na qual o documento serve regularmente para a administração; idade semiativa, quando o documento é utilizado ocasionalmente, e, na idade inativa, é destruído ou encaminhado para a preservação permanente (ROUSSEAU; COUTURE, 1998).

O Dicionário Brasileiro de Terminologia Arquivística define a teoria das três idades como a “Teoria segundo a qual os arquivos são considerados arquivos correntes, intermediários ou permanentes, de acordo com a frequência de uso por suas entidades produtoras e a identificação de seus valores primários e secundários” (ARQUIVO NACIONAL, 2005, p. 160).

Os princípios localizados na base da teoria arquivística, segundo Bellotto (2006), são:

- 1 *Princípio da proveniência*: fixa o documento em relação a seu produtor. Estabelece que arquivos originários de uma instituição ou de uma pessoa não devem ser mesclados, no arquivo, a outros com origem distinta;
- 2 *Princípio da organicidade*: “A organicidade é a qualidade segundo a qual os arquivos espelham a estrutura, as funções e as atividades da entidade produtora/acumuladora em suas relações internas e externas”;
- 3 *Princípio da unicidade*: estabelece que “os documentos de arquivo conservam seu caráter único, em função de seu contexto de produção”;
- 4 *Princípio da indivisibilidade ou integridade administrativa*: “os fundos de arquivos devem ser preservados sem dispersão, mutilação, alienação, destruição não autorizada ou adição indevida”;
- 5 *Princípio da cumulatividade*: “o arquivo é uma formação progressiva, natural e orgânica”.

O princípio da proveniência é também chamado de princípio de respeito aos fundos, portanto, o fundo é um conjunto de documento de mesma proveniência. (ARQUIVO NACIONAL, 2005).

O princípio da proveniência é de tal forma basilar na Arquivologia que Rousseau e Couture (1998, p. 95) afirmam que:

Não se pode, pois, desempenhar tarefa de arquivista sem recorrer ao princípio da proveniência, sem o qual não pode haver fundo de arquivo. Finalmente, estamos agora aptos a julgar da necessidade incontestável do princípio da proveniência tanto no plano da organização e do tratamento dos arquivos para fins administrativos como no da sua organização e tratamento para fins patrimoniais ou de investigação.

Rousseau e Couture (1998) levantam, ainda, que a questão informação eletrônica tem levado ao raciocínio de que a noção de fundo corre riscos, pela impossibilidade de precisar a origem de uma informação eletrônica, e que, por consequência, o princípio da proveniência não poderia continuar a existir. É uma questão extremamente polêmica, considerando que, no princípio da proveniência, a concepção de fundo e a organicidade fazem parte da base da arquivística.

Muitos são os estudiosos que defendem a necessidade de repensar a arquivística no contexto da informatização da sociedade.

Segundo Cook (2001), é necessário repensar as ciências arquivísticas, que possuem raízes no Positivismo do século XIX, com conceitos, estratégias e metodologias que não são mais viáveis em um mundo pós-moderno e informatizado. Para ele, no coração do novo paradigma, está a mudança da visão de documentos arquivísticos como objetos físicos estáticos, para o entendimento deles como conceito virtual dinâmico, isto é, o deixar de ver os documentos arquivísticos como produtos passivos da atividade administrativa, e passar a considerá-los como agentes ativos na formação da memória de organizações e pessoas.

A simbiose entre a informação e a tecnologia marcou a entrada da arquivística na era pós-custodial (RIBEIRO, 2001).

Uma teoria considerada pós-custodial é o *Record Continuum*, proposto por Atherton (1985), adotado pela Austrália. A custódia, para o arquivista pós-custodial, é apenas uma das muitas questões que devem ser enfrentadas nas múltiplas realidades dos arquivos virtuais (UPWARD, 1996).

A concepção *continuum* desafia os entendimentos tradicionais que diferencia arquivos de documentos arquivísticos com base a seleção para preservação permanente na custódia de um arquivo e com foco na sua natureza fixa, com adoção de uma visão pluralista da informação registrada, caracterizando documentos arquivísticos como um gênero de

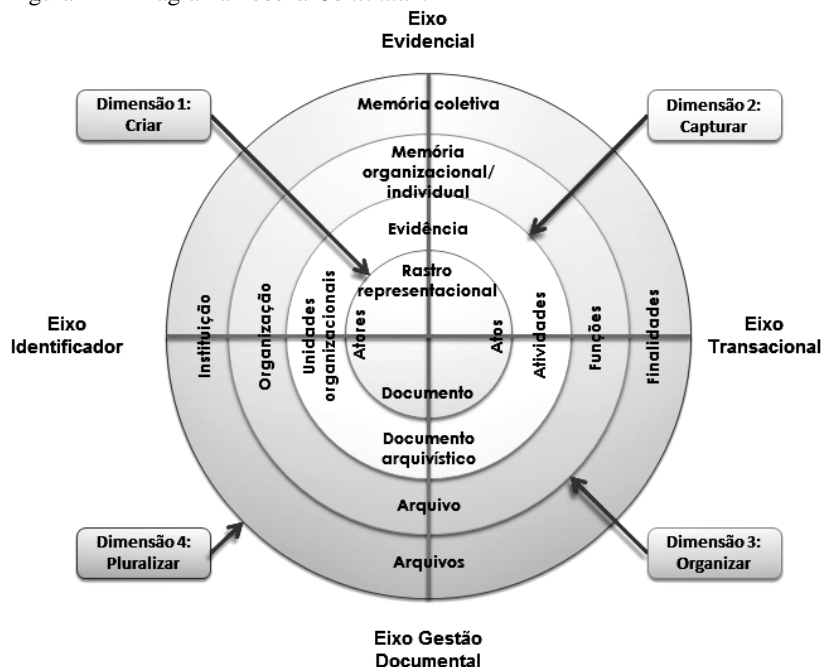
documentos em termos da sua intenção e funcionalidade. Ele enfatiza a natureza probatória, transacional e contextual, rejeitando abordagens do documento arquivístico com o foco no conteúdo e valor informacional (MCKEMMISH, 2001).

O modelo *continuum* considera a natureza do documento como fixo e mutável ao mesmo tempo, enquanto o conteúdo e a estrutura do documento podem ser vistos como fixos, seu contexto é sempre mutável (MUGICA, 2005).

Os quatro princípios base do modelo *record continuum* são (UPWARD, 1996):

- 1 um conceito de documento arquivístico, inclusive dos documentos de valor permanente (arquivos), ressalta seus usos para o propósito transacional, de evidência e de memória, unificando abordagens para arquivamento e gestão, ainda que os documentos sejam mantidos por uma fração de segundos ou um milênio;
- 2 os documentos arquivísticos têm foco como entidades lógicas em vez de físicas, independente deles estarem em papel ou em formato digital.
- 3 a institucionalização do papel do profissional de gestão de documentos arquivísticos requer uma ênfase na necessidade de integrá-la aos propósitos e processos sociais e de negócios; e
- 4 a ciência arquivística é a base para organizar o conhecimento sobre gestão de documentos arquivísticos.

Upward (1996) apresenta as propriedades do *record continuum* em quatro eixos e quatro dimensões (Figura 21).

Figura 21 – Diagrama *Record Continuum*

Fonte: Adaptado de Upward (1996).

Nos quatro eixos, estão presentes os principais temas da arquivística, cada eixo com quatro coordenadas que podem ter ligações dimensionais:

- O eixo gestão documental lida com os veículos para o armazenamento de informações registradas sobre atividades humanas. No modelo, o documento é uma pseudorrepresentação de um ato, com conteúdo, estrutura e contexto de criação. Este eixo conta, ainda, com documentos arquivísticos (que possuem ligações com outros documentos), o arquivo que é o agregado de documentos arquivísticos na visão da sua totalidade numa organização, e os arquivos na forma plural que contêm os documentos arquivísticos de um conjunto de organizações;
- O eixo evidencial consiste nos rastros de ações. São as evidências que os documentos arquivísticos podem fornecer, e seu papel na memória corporativa e coletiva;

- O eixo transacional apresenta os atos, as atividades, as funções e as finalidades como coordenadas. Este eixo reflete a ênfase por documentos, como documentos arquivísticos de atividades realizadas na condução de transações (negócios) e na forma como essas atividades criam ligações entre documentos¹²;
- O eixo identificador representa o ator, a unidade de trabalho na qual o ator está associado, a organização na qual a unidade de trabalho está associada e a forma pela qual as identidades desses elementos são institucionalizados por amplo reconhecimento social. É o eixo de proveniência estrutural de autoridades e responsabilidades dentro do qual os arquivos são feitos e usados. É responsável por um dos principais temas da arquivística, na qual um arquivo deve estar ligado ao criador de documentos, e que os documentos refletem a autoridade e responsabilidade que dá suporte a um ato.

As quatro dimensões são reveladas com as conexões das coordenadas dos eixos, no entanto, Upward (1996) enfatiza que as dimensões não são fronteiras e as coordenadas não estão presentes sempre. As coisas podem acontecer simultaneamente nas dimensões, não importando como o sistema de gestão documental está criado.

A seguinte leitura das dimensões do modelo *continuum* é apresentada por McKemmish (1997):

- 1a dimensão) Criar – engloba os atores que realizam o ato, os atos em si, os documentos que registram os atos e o rastro representacional do ato;
- 2a dimensão) Capturar – engloba os sistemas de gestão de arquivos pessoais e corporativos que capturam documentos como evidências das atividades sociais e de negócios;
- 3a dimensão) Organizar – abrange a organização dos processos de gestão de documentos, visando à forma como uma corporação ou indivíduo define o seu regime de gestão de documentos, e a formação do arquivo como memória das suas funções sociais e de negócios;

¹² Na arquivística, essas ligações entre documentos são conhecidas como relações orgânicas que proveem a organicidade entre os documentos.

4a dimensão) Pluralizar – diz respeito à forma pela qual os arquivos são colocados em uma estrutura a fim de prover uma memória coletiva, histórica e cultural dos propósitos sociais institucionalizados.

A visão de mundo do *record continuum* é distinta do ciclo de vida. Segundo McKemmish (1997), na visão do *continuum*:

- os documentos arquivísticos, embora congelados no tempo, fixos em uma forma documental e ligados ao seu contexto de criação, são, ao mesmo tempo, atuais e históricos desde a sua criação;
- as dimensões não são baseadas no tempo, porém, os documentos são acoplados ao espaço e tempo, e perpetuamente conectados a eventos no passado;
- o foco está no documento arquivístico lógico e seu relacionamento com outros documentos, e os contextos de criação e uso;
- o modelo é um mapa de um local dinâmico e virtual, com múltiplas realidades, em que os documentos estão sem amarras e são transportados por meio do tempo e do espaço, reapresentados no contexto do seu uso.

Segundo Duranti (2005), o conceito de *record continuum* foi popularizado na Austrália, numa ótica em que os documentos são produzidos e mantidos pelo produtor ou conservador, independente da manutenção durar uma fração de segundo ou permanentemente, as atividades são as mesmas. Salienta, ainda, que a comunidade internacional e o Conselho Internacional de Arquivos voltam ao ciclo de vida para documentos eletrônicos.

2.4.2 Avaliação

Na Arquivologia, a avaliação consiste, fundamentalmente, em identificar valores de documentos com vistas a estabelecer prazos para guarda ou eliminação (BERNARDES, 1998).

O acúmulo de documentos torna a análise de documentos bastante trabalhosa, levando, com frequência, à eliminação sem critério ou, à reprodução do acervo em outros suportes, sem a prévia identificação e avaliação dos conjuntos documentais (BERNARDES, 1998).

A preocupação de qual documento deveria ser preservado já havia sido manifestada no clássico artigo de Brooks (1940) *The*

selection of records for preservation, que tem início com a pergunta “*What records shall we preserve?*”. Embora a realidade da época se limitasse aos documentos tradicionais (em papel), várias considerações continuam aplicáveis na realidade digital.

Brooks defende que, idealmente, a seleção deve ser contínua, com a eliminação de cópias de documentos, excetuando aqueles que se tornam documentos formais.

Brooks estabelece três categorias de valores para os documentos arquivísticos: a primeira relacionada ao valor para agência criadora do documento, diz respeito à eficiência administrativa e proteção contra reclamações de todo; a segunda categoria seria o valor como fonte de estudo da história administrativa da agência criadora do documento; e a terceira categoria seria o valor histórico (BROOKS, 1940; PINKETT, 1981).

Quanto ao conhecimento necessário para avaliação e seleção de documentos arquivísticos, Brooks (1940) identifica quatro tipos: primeiro, saber a origem, história, objetivo e métodos da agência; segundo, conhecer os relacionamentos entre os documentos arquivísticos; terceiro, saber e estar atento às mudanças no escopo e nos métodos de pesquisa; e quarto, estar familiarizado com o uso que se faz com os documentos.

Schellenberg (2006) justifica a necessidade da avaliação dos documentos públicos na importância da redução da quantidade de documentos, tanto para o governo quanto para os pesquisadores. A abordagem schellenberguiana tem base nos valores dos documentos. Nela, é fixada a distinção entre os valores primários e secundários. O primeiro, para a própria entidade criadora/receptora do documento, e a segunda, para outras entidades e utilizadores. O valor secundário é a utilidade permanente do documento e pode ser determinado com a análise de dois aspectos: a) o valor probatório do documento; e b) o valor informativo do documento (SCHELLENBERG, 2006).

Schellenberg (2006) estabelece critérios para testar os valores probatórios e informativos dos documentos.

Para efetuar os testes para os valores probatórios, Schellenberg estabelece como requisito o conhecimento completo da documentação do órgão e o conhecimento do significado de grupos de documentos produzidos em relação às suas funções de maior importância. A fim de auxiliar o processo de análise, apresenta onze questões agrupadas sob três aspectos da organização (Quadro 13).

Quadro 13 – Questões para análise do valor probatório

Aspecto	Questão
Estrutura	<ol style="list-style-type: none"> 1. Que unidades do órgão têm responsabilidade primordial pela formulação de decisões concernentes a sua organização, seus programas e suas normas? 2. Que unidades exercem atividades auxiliares à formulação de tais decisões? 3. Que funcionários, fora da sede, têm arbítrio para formular tais decisões? 4. Que séries de documentos são essenciais para refletir sobre tais decisões?
Funções	<ol style="list-style-type: none"> 5. A que funções do órgão os documentos se relacionam? 6. São as funções substantivas? 7. Que série de documentos é essencial para mostrar como cada função substantiva foi levada a efeito em cada nível da organização, tanto no serviço central como nos locais?
Processos/ atividades	<ol style="list-style-type: none"> 8. Que atividades de supervisão e chefia estão envolvidas na administração de cada função? 9. Quais são as operações sucessivas em sua execução? 10. Que documentos dizem respeito à direção executiva, distintos da execução da função? 11. Que documentos devem ser preservados, a título de exemplo, para mostrar os processos de trabalho nos níveis inferiores da organização?

Fonte: Baseado em Schellenberg (2006).

Na visão de Schellenberg (2006), na apuração do valor informativo dos documentos, a origem – o produtor e a atividade – não são importantes. O que importa é o conteúdo, que pode ser analisado sob três fatores: a) unicidade; b) forma; e c) importância (Quadro 14).

Quadro 14 – Fatores para teste de valor informativo

Fator	Característica
Unicidade	<p>Informação não encontrada em outras fontes documentárias de forma tão completa e utilizável.</p> <p>Todas as outras fontes de informação sobre o assunto devem ser examinadas.</p> <p>Para determinar a unicidade do documento, o avaliador deve ser um perito no conteúdo, familiarizado com os recursos externos relacionados.</p> <p>Considerar a duplicidade física do documento.</p>

Forma	<p>Considerar a forma da informação nos documentos e a forma dos documentos.</p> <p>A forma da informação refere-se, principalmente, ao grau de concentração da informação: a) poucos fatos apresentados sobre muitas pessoas, coisas ou fenômenos – informação extensa; b) muitos fatos sobre algumas pessoas, coisas ou fenômenos – informação intensa; ou c) muitos fatos sobre diversas matérias – informação diversa.</p> <p>Em geral, os documentos que representam maior concentração de informação são as mais adequadas para preservação arquivística.</p> <p>A forma de documentos refere-se à condição física dos documentos.</p>
Importância	<p>Prioridade às necessidades do historiador, cientista social e genealogista.</p> <p>Preservar também para usuários de campos técnicos especializados e cientistas que não fazem amplo uso de documentos no exercício normal de suas profissões.</p> <p>Os documentos poder ter tanto significação coletiva quanto individual. O valor da pesquisa, geralmente, deriva da importância da informação num conjunto de documentos e não em peças isoladas.</p> <p>O valor do documento normalmente está relacionado com a importância da pessoa ou coisa.</p> <p>Documentos podem ter valores sentimentais.</p> <p>Resultados positivos dos testes de Unicidade e Forma antecedem o teste Importância.</p> <p>O teste da importância refere-se a fatores imponderáveis, enquanto os testes de unicidade e forma relacionam-se a fatores ponderáveis.</p>

Fonte: Baseado em Schellenberg (2006).

Os testes para valores informativos são aplicados em grupos de documentos identificando informações relativas a três elementos: a) pessoas; b) coisa; ou c) fatos (SCHELLENBERG, 2006).

Na estratégia de analisar o valor dos documentos, Terry Cook (1995, 2001b, 2006) concebe a teoria e a prática da macroavaliação (*macroappraisal*), adotada pelo Arquivo Nacional do Canadá.

Na concepção de Cook (2006), a macroavaliação analisa o valor da sociedade no contexto funcional e estrutural, e a cultura local, na qual os documentos arquivísticos são criados.

A avaliação tem foco no valor de longo prazo do conteúdo dos documentos arquivísticos, por seus valores potenciais de pesquisa, a

macroavaliação analisa a importância do contexto de sua criação e utilização.

2.4.2.1 Operação Metodológica de Silva e Ribeiro

Silva e Ribeiro (2000, 2004, 2009) propõem a avaliação arquivística como uma operação metodológica e não simplesmente um procedimento de carácter prático. Para o desenvolvimento da proposta, os autores assumem a perspectiva da informação (e não o documento) como objeto de estudo e trabalho.

Partindo da definição de Informação como conjunto estruturado de representações mentais codificadas (símbolos significantes) socialmente contextualizadas e passíveis de serem registadas num qualquer suporte material (papel, filme, banda magnética, disco compacto, etc.) e, portanto, comunicadas de forma assíncrona e multidirecionada, temos, desde logo, que passar a valorizar determinados aspectos relativos à própria informação que, até agora, numa lógica documentalista, têm sido desprezados ou mesmo ignorados (SILVA; RIBEIRO, 2009).

A operação metodológica proposta por Silva e Ribeiro (2000, 2004, 2009) é assentada em três parâmetros e critérios de avaliação e suas combinações:

- **Pertinência:** que significa pertencer à ação de indivíduo ou entidade, mensurado por meio do trinômio objetivos essenciais + estrutura orgânica e competências/funções + memória. Recebe três níveis de gradação: relação direta (A), indireta (B) ou periférica (C);
- **Densidade:** que significa a qualidade da densidade informacional. Avalia se o documento é primário/original ou se é secundário (informação derivada da original) e se possui cópia/duplicação;
- **Frequência:** significa a quantificação da periodicidade de uso/acesso da informação, quer na fase de produção ou recepção, quer na fase imediatamente posterior, perene e definitiva.

A decisão sobre o destino final da informação, na metodologia de Silva e Ribeiro, tem base no Quadro 15.

Quadro 15 – Decisão sobre o destino da informação

Pertinência	Densidade	Frequência	Destino Final
Nível A	1	1 ou 0	Conservação Permanente

Nível A	0	1	Conservação Temporária
Nível A	0	0	Eliminação
Nível B	1	1 ou 0	Conservação Permanente
Nível B	0	1	Conservação Temporária
Nível B	0	0	Eliminação
Nível C	1 ou 0	1	Conservação Temporária
Nível C	1 ou 0	0	Eliminação

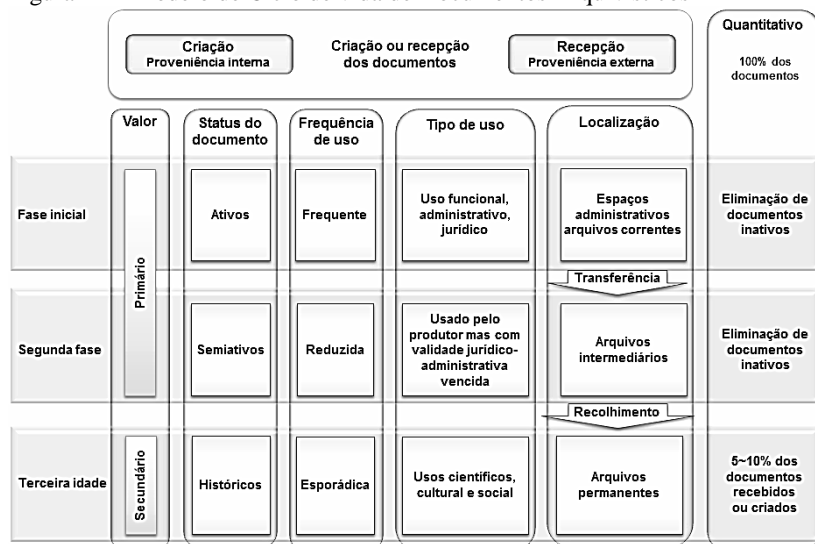
Fonte: Adaptado de Silva e Ribeiro (2000, 2004, 2009).

2.4.3 Conhecimento da Arquivística na preservação do conhecimento

Dos dois modelos que dão origem aos processos de gestão de documentos arquivísticos, embora o *record continuum* tenha surgido como uma alternativa a abordagem das três idades (ciclo de vida), para a preservação do conhecimento codificado armazenado em objetos digitais, mostrou-se adequado como fundamento, tanto pela adequabilidade para o contexto dos objetos digitais, tanto pela questão cultural da Arquivologia no Brasil.

A Figura 22 fornece uma visão resumida do modelo, explicitando seis características do documento no ciclo de vida: valor, *status*, frequência de uso, tipo de uso, localização e quantitativo.

Figura 22 – Modelo de Ciclo de vida de Documentos Arquivísticos



Fonte: Baseado em Rousseau e Couture (1998) e Bellotto (2006).

O modelo do ciclo de vida criado na arquivística foi relido e adaptado no contexto do objeto digital com o conhecimento codificado.

É possível raciocinar que, analogamente aos documentos arquivísticos, o uso e acesso ao conhecimento armazenado em objetos digitais deve decrescer à medida do seu envelhecimento.

Porém, como unicamente a idade do objeto não pode ser determinante, outros critérios, como o tipo de uso e a frequência, devem ser utilizados na decisão de quais e como os objetos digitais devem ser preservados.

As estimativas do percentual de documentos arquivísticos (criados e recebidos) que são selecionados para a preservação permanente variam. Rousseau e Couture (1998) estimam em 5% a 10% e, de acordo com *Records Management Unit* do governo dos Territórios do Norte do Canadá, somente 3% a 5% dos documentos arquivísticos são preservados, isto é, cerca de 95% a 97% são eliminados (GOVERNMENT OF THE NORTHWEST TERRITORIES, 2002).

Os arquivos existem há séculos de várias formas, mas os princípios da arquivística foram articulados em detalhes, no século XIX, na França e na Alemanha (COOK, 1997). Se, numa área de conhecimento com mais de um século, as estimativas são ainda imprecisas, certamente, no tocante ao percentual de objetos digitais que

devem ser preservados com vistas à preservação do conhecimento, a incerteza é muito maior.

Similarmente à arquivística, outro fator que influencia significativamente o volume de objetos digitais são as cópias. Na realidade da primeira metade do século passado, Brooks (1940) já afirmava que a quantidade de duplicação era enorme e que poderia chegar à metade dos documentos arquivísticos do governo.

A avaliação de documentos arquivísticos pode servir de fonte de aprendizado, porém, o foco é distinto – do documento arquivístico ao conhecimento codificado – e a realidade da sociedade altamente informatizada e conectada é diferente daquela em que uma significativa parte do arcabouço de conhecimento relativo à avaliação arquivística foi desenvolvida.

Schellenberg (2006), na década de 50 do século passado, já afirmava que à proporção que outras espécies de materiais documentários aumentam, diminui a quantidade de documentos oficiais que merecem preservação permanente, logo, é duvidoso que os governos possam justificar a guarda de mais do que uma pequena proporção de documentos oficiais. Afirma, ainda, que a tarefa de avaliação se torna mais complexa à medida que a documentação da sociedade aumenta.

Para a avaliação da importância do documento, Schellenberg (2006) argumenta que o conhecimento especializado do arquivista é adquirido na experiência com os documentos com que trabalha e, ao executar o serviço de referência, toma conhecimento das necessidades de pesquisa. Adquire, também, o conhecimento sobre a documentação produzida pelo órgão, o que permite reduzi-la a proporções manuseáveis para ser usada em pesquisas.

As definições e características de documentos arquivísticos apontadas por Thomassen (2001) são características fundamentadas que podem se transformar em critérios para seleção de objetos digitais do conhecimento. Para ele, documentos arquivísticos são:

- distintos de outros documentos pelo motivo da sua criação;
- são informações ligadas a processos de trabalho, que pode ser processo-chave (relacionado com os objetivos da organização) ou processo de apoio;
- funcionam como memória de indivíduos, de organizações e da sociedade;
- servem como agentes de controle e evidência.

2.5 PRESERVAÇÃO DIGITAL

A preocupação em preservar é antiga, mas a preservação digital tem diferenças fundamentais em relação à preservação de outros objetos, como o arqueológico, histórico, documentos em papel, entre outros.

Na preservação do documento em papel, a preservação do suporte (o papel) resulta na preservação do conteúdo, em caso de danos, frequentemente, as perdas são parciais. No caso do objeto digital, a preservação do suporte é somente uma pequena parte do processo de preservação, e com tendência de ser cada vez menos importante com a evolução da tecnologia de comunicação e armazenamento, como a computação em nuvem.

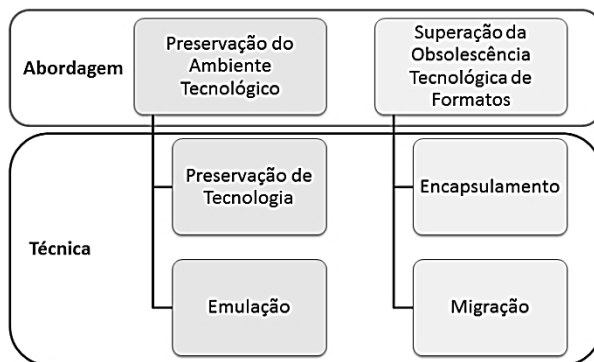
2.5.1 Abordagens e técnicas de Preservação Digital

A preservação digital consiste na manutenção da capacidade de apresentar os elementos essenciais dos objetos digitais e deve fazer frente a perigos que possam ameaçar.

O patrimônio digital sofre dois tipos de ameaças: a física e a técnica. A física é referente aos riscos de danos na mídia e resultantes de incidentes climáticos, e a técnica é a dificuldade ou incapacidade de acessar e usar os objetos digitais devido à evolução técnica do *hardware* e do *software* (CONSTANTOPOULOS; DRITSOU, 2007). Neste sentido, instituições de pesquisas e empresas investem no desenvolvimento de técnicas a fim de reduzir esses riscos.

As técnicas de preservação digital podem ser classificadas em duas principais abordagens: 1) preservação do ambiente tecnológico; e 2) superação da obsolescência tecnológica de formatos (Figura 23).

Figura 23 – Técnicas de Preservação Digital



Fonte: Adaptado de Lee *et al.* (2002).

Um resumo dessas técnicas é apresentado no Quadro 16.

Quadro 16 – Técnicas de Preservação Digital

Abordagem	Técnica	
Preservação do ambiente tecnológico	Preservação de tecnologia	Manter um museu completo de equipamentos obsoletos para a replicação de qualquer velha configuração. Implica na preservação de <i>software</i> (sistemas operacionais e aplicativos) e <i>hardware</i> . É uma solução de curto prazo, é complexo e envolve altos custos para manutenção e provimento de espaço físico.
	Emulação	É a emulação de plataformas e sistemas operacionais obsoletos em novos sistemas, preservando os programas aplicativos originais. Visa preservar a aparência e as funcionalidades dos objetos digitais. Técnica com defensores e críticos. Requer preservar as especificações detalhadas dos <i>hardwares</i> e sistemas operacionais obsoletos.
Superação da obsolescência tecnológica de formatos	Migração	Estabelece a transferência periódica do objeto digital de uma plataforma tecnológica para outra mais nova, ou de uma geração de tecnologia para uma geração subsequente, com objetivo de preservar a integridade dos objetos

		digitais e manter a habilidade para usuários recuperar, visualizar e usar.
	Encapsulamento	O objeto digital e todas as informações necessárias para prover o acesso a ele devem ser agrupadas e encapsuladas em container, que é estrutura lógica ou física.

Fonte: Baseado em Lee *et al.* (2002).

A abordagem da preservação do ambiente tecnológico é mais conservadora e o ambiente tecnológico original deve ser totalmente preservado para permitir que os objetos digitais gerados nesse ambiente possam ser decodificados no futuro. A segunda abordagem busca superar o problema da obsolescência tecnológica dos formatos dos objetos digitais.

A migração é, atualmente, a estratégia mais utilizada (SAYÃO, 2010) e mais aceita, mas também apresenta problemas, como o alto custo e a perda de dados tanto na gravação como da incompatibilidade de formato (SERRA, 2001), confirmando a complexidade do problema da preservação digital. Todas as técnicas apresentadas trazem vantagens, desvantagens e riscos.

2.5.2 Abordagem Sociotécnica de Preservação

A abordagem sociotécnica tem como principais autores Frederic Emery e Eric Trist, nessa, consideram que as empresas são compostas pelo subsistema social e o subsistema técnico, interdependentes e interativos (MARTINS, 2006).

A visão sociotécnica vem para suprir as carências e dificuldades das abordagens fundamentalmente centradas somente no subsistema técnico, com a adoção do tripé pessoa-processo-tecnologia.

Na preservação digital, a atuação do subsistema sócio junto com o subsistema técnico e um processo bem definido, é vital.

Na visão desenvolvida pelo projeto InterPARES1 (DURANTI; SUDERMAN; TODD, 2008), o subsistema sócio é subdivido em dois papéis: o produtor (*creator*) e o preservador (*preserver*). Em oposição à concepção que a preocupação da preservação é inteiramente atribuída ao preservador, o produtor tem importante participação no processo, isto é, a preservação tem início na produção do objeto digital.

Nos guias desenvolvidos pelo InterPARES, observa-se que foram estabelecidas mais diretrizes para o produtor do que para o preservador.

Excertos das Diretrizes para o Preservador (*Preserver Guidelines*) (INTERPARES PROJECT, 2007) e das Diretrizes para o Produtor (*Creator Guidelines*) (INTERPARES PROJECT, 2007) são apresentados no Quadro 17 e no Quadro 18, respectivamente.

Quadro 17 – Diretrizes para o Preservador

Item	Diretriz
Gestão	Gerenciar a cadeia do <i>framework</i> de preservação.
Avaliação	Avaliar documentos para preservação permanente.
Recepção	Receber documentos selecionados para preservação permanente.
Preservação	Preservar os documentos recebidos.
Saída	Dar acesso aos documentos.

Fonte: Baseado em InterPARES Project (2007a).

Quadro 18 – Diretrizes para o Produtor

Item	Diretriz
Acessibilidade	Selecionar <i>hardwares</i> , <i>softwares</i> e formatos de arquivo que ofereçam as melhores expectativas de que permanecerão acessíveis ao longo do tempo.
Fixidez	Certificar que os materiais digitais são estáveis no conteúdo e na forma.
Identidade	Certificar que os materiais digitais estão identificados adequadamente.
Integridade	Certificar que os materiais digitais carregam informações que ajudam a verificar a sua integridade.
Organização	Agrupar os materiais de forma lógica.
Autenticação	Utilizar técnicas de autenticação que favoreçam a manutenção e preservação.
Proteção	Proteger os materiais digitais de ações não autorizadas.
Cópia de segurança	Proteger os materiais digitais de perdas acidentais e corrupção.
Obsolescência	Prevenir contra a obsolescência de <i>software</i> e <i>hardware</i> .
Consciência	Considerar os aspectos relacionados à preservação em longo prazo.

Fonte: Baseado em InterPARES Project (2007b).

No processo tradicional de preservação, um ponto comum é manter o objeto inalterado para conservá-lo o mais próximo possível do estado original. No caso da preservação digital, a alteração é inevitável pela necessidade da permanente atualização tecnológica do objeto digital a fim de mantê-lo acessível com o uso das tecnologias

disponíveis, ou seja, atualizadas – é o paradoxo da preservação digital (CHEN, 2001; THIBODEAU, 2002).

Garantir a acessibilidade permanente do patrimônio digital durante o tempo necessário é o objetivo da Preservação Digital, devendo, para isso, encontrar formas (processos, *hardware* e *software*) de representar o que se havia apresentado originalmente aos usuários (NATIONAL LIBRARY OF AUSTRALIA, 2003).

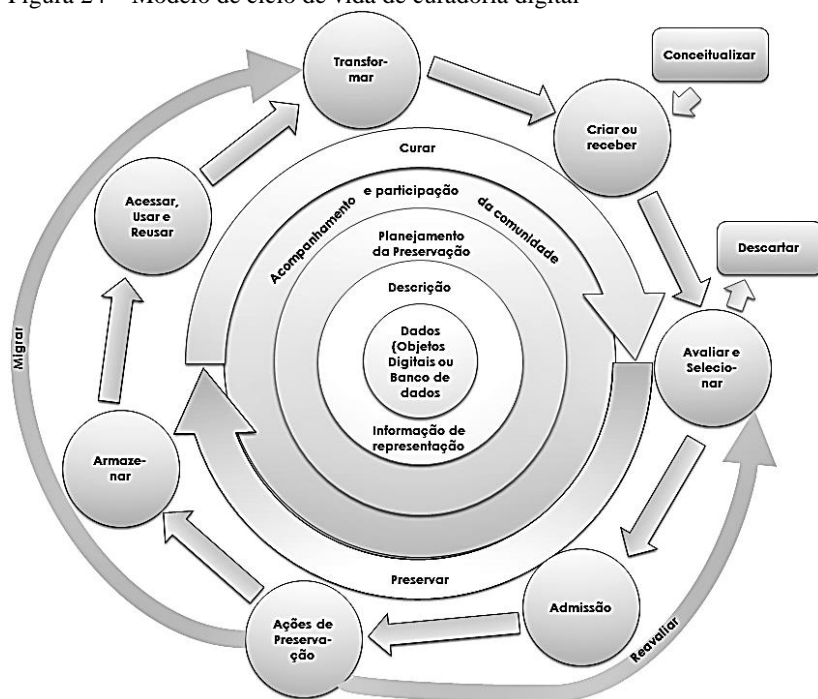
2.6 CURADORIA DIGITAL

A Curadoria Digital é a avaliação e a gestão ativa dos dados digitais ao longo do seu ciclo de vida, em que se consideram os processos para a manutenção, preservação e agregação de valor aos dados (DIGITAL CURATION CENTRE, [s.d.]).

O objetivo de um programa de Curadoria Digital é salvaguardar os objetos digitais, possibilitando o acesso e o reuso em todo seu ciclo de vida. O retorno do investimento da curadoria vem do compartilhamento dos dados, evitando a repetição de esforços na criação destes, tornando-os disponíveis para extração de novos conhecimentos. Dessa forma, a curadoria digital permite:

- a) manter o documento íntegro e acessível, enquanto este possuir valor jurídico (evidência);
- b) extrair novos conhecimentos (valor informacional e de pesquisa);
- c) preservar a memória da sociedade (valor histórico); e
- d) evitar o retrabalho de recriar os dados já produzidos anteriormente.

Figura 24 – Modelo de ciclo de vida de curadoria digital



Fonte: Yamaoka (2012).

O *Digital Curation Centre* (DCC) concebeu um Modelo de Ciclo de Vida de Curadoria (Figura 24) que contempla as etapas necessárias para a curadoria de dados. Essa abordagem visa garantir que todas as etapas requeridas sejam identificadas e planejadas, e que as ações sejam encadeadas numa sequência correta, o que garante a manutenção da autenticidade, confiabilidade, integridade e usabilidade do objeto digital (HIGGINS, 2008).

Os elementos que compõem o modelo de ciclo de vida são os dados, as ações contínuas, as ações sequenciais e as ações ocasionais (HIGGINS, 2008).

O elemento “dados” representa qualquer informação digital binária, é o centro do ciclo de vida de curadoria. Aquele pode ser os objetos digitais simples, complexos, e o banco de dados.

As “ações contínuas” são as ações que abrangem todo o ciclo de vida. São elas (HIGGINS, 2008):

1. ***descrição e representação da informação***: atribuição dos metadados (administrativo, descritivo, técnico, estrutural e de preservação) para garantir descrição e controle no tempo. Inclui a coleta e atribuição da informação de representação requerida para entendimento e “renderização” do objeto digital e dos metadados associados;
2. ***planejamento da preservação***: é o plano para preservação em todo o ciclo de vida do objeto digital, incluindo a gestão de todas as ações de curadoria;
3. ***acompanhamento e participação da comunidade***: é o processo de participação no desenvolvimento de padrões comuns, ferramentas e *software*; e
4. ***curadoria e preservação***: é encaminhar a gestão e as ações planejadas para promover a curadoria e preservação.

As “ações sequenciais” são aquelas que devem ser executadas ciclicamente (HIGGINS, 2008):

1. ***conceitualização***: compreende conceber e planejar a criação de dados, incluindo o método de captura e opções de armazenamento;
2. ***criação ou recebimento***: para criar dados, incluir metadados administrativos, descritivo, técnico, estrutural e de preservação; ou receber dados dos criadores, das instituições de arquivos, dos repositórios ou dos centros de dados e, se necessário, atribuir metadados;
3. ***avaliação e seleção***: ação que visa avaliar e selecionar os dados para curadoria e preservação de longo prazo, que deve ser aderente às orientações documentadas, políticas e exigências legais;
4. ***absorção (ingest)***: para destinar os dados para um arquivo, repositório, data center ou outra entidade de custódia;
5. ***ações de preservação***: visa assegurar a integridade dos dados (autênticos, confiáveis e usáveis), com limpeza de dados, validação, atribuição de metadados de preservação, informações representação e garantir estruturas de dados ou formatos de arquivos aceitáveis;
6. ***armazenamento***: visa armazenar dados de forma segura e aderente aos padrões relevantes;
7. ***acesso, uso e reuso***: para garantir o acesso com um robusto controle de acesso e de autenticação dos usuários;
8. ***transformação***: visa à criação de novos dados a partir do original, como, por exemplo, a migração para um formato

diferente ou a geração de resultados derivados por seleção ou consulta.

As “**ações ocasionais**” são aquelas aplicadas eventualmente, decorrente de uma decisão, pode reordenar as ações sequenciais. São elas (HIGGINS, 2008):

1. **eliminação**: os dados que não foram selecionados para curadoria e preservação podem ser descartados, transferidos para outro arquivo, depósito, data center ou outro centro de custódia;
2. **reavaliação e seleção**: visa à nova avaliação e seleção de dados retornados por falhas nos procedimentos de validação;
3. **migração**: migrar o dado para um formato diferente para garantir a imunidade da obsolescência de *hardware* e *software*.

Em síntese, a Curadoria Digital é a gestão e preservação de longo prazo dos dados digitais (ABBOTT, 2008) que surge como uma nova área de prática e de pesquisa que faz uso do conhecimento de várias disciplinas (SAYÃO; SALES, 2012).

2.7 OBJETO DIGITAL

É requisito em todo processo de preservação, o conhecimento do objeto a ser preservado, como a sua composição, características e fragilidades. Esta seção trata do elemento central deste trabalho, o **objeto digital**, descrevendo a busca por uma conceituação que seja aderente ao objetivo estabelecido.

Uma das maiores dificuldades em uma disciplina emergente como a preservação digital é a falta da precisão terminológica, consequentemente, a comunicação pode tornar-se problemática se as comunidades usarem o mesmo termo com significados diferentes (JONES; BEAGRIE, 2008), e termos diferentes com o mesmo significado.

A preservação digital ainda não alcançou uma teoria consolidada, nessa busca, é justificável refinar os modelos e as terminologias de suporte do assunto (LUDWIG, 2010).

Sem a clareza dos conceitos acerca daquilo que será discutido, o potencial para confusão é muito grande (HARVEY, 2005).

O termo objeto digital é o mais ambíguo, uma vez que se refere tanto aos aspectos conceituais e técnicas, ou seja, o conteúdo intelectual e sua forma, quanto à forma como é representada digitalmente (HOFMAN, 2002).

Com frequência, os termos objeto digital, arquivo de computador (*file*), documento digital, artefato digital, recurso digital e material digital são utilizados como sinônimos.

Para englobar os diferentes elementos que constituem o patrimônio digital, o termo **material digital** é adotado em diversos trabalhos, entre eles: no *Guidelines for the Preservation of Digital Heritage* da UNESCO (NATIONAL LIBRARY OF AUSTRALIA, 2003); no *Preservation Management of Digital Materials – The Handbook* (JONES; BEAGRIE, 2008); e, também, no livro *Preserving Digital Material* (HARVEY, 2005).

O *Handbook* de Jones e Beagrie (2008) e o *Guidelines da National Library of Australia* (2003) usam também o termo objeto digital, sem fazer distinção explícita do termo material digital. Harvey (2005), por sua vez, adota o termo material digital em concordância com o *Handbook* e o *Guideline*, porém, faz clara distinção dos termos objeto digital e artefato digital, que têm os seguintes conceitos: artefato digital é a mídia de armazenamento, isto é, o suporte mais as cadeias de bits gravadas nele; e o objeto digital é formado pelas cadeias de bits mais todas as coisas necessárias para dar sentido a essas cadeias.

Esses conceitos mais gerais contribuem pouco na elaboração de modelos e arquiteturas de preservação de objetos digitais. A proposta de Thibodeau (2002) de conceituar objetos digitais em três camadas possibilita identificar com mais clareza o objeto a ser preservado.

2.7.1 As classes de objetos digitais de Thibodeau

Objeto digital é um objeto de informação, de qualquer tipo e formato, expresso sob a forma digital (THIBODEAU, 2002). Dessa definição geral, Thibodeau (2002) propõe que os objetos digitais herdem as propriedades de três classes:

- a) **objeto físico** – como objeto físico, o objeto digital é simplesmente uma inscrição de sinais em uma mídia. O meio físico determina uma convenção para a gravação de dados com densidade e tamanho de blocos diferentes. A inscrição física é independente do significado e, portanto, o computador não sabe se o objeto contém um documento em linguagem natural ou uma foto, por exemplo. A inscrição física não implica em sintaxe, morfologia ou semântica;
- b) **objeto lógico** – como objeto lógico, o objeto digital é reconhecido e processado por *software*. No nível lógico, a gramática é independente da inscrição física. Um *software*

aplicativo reconhece o formato do objeto, os tipos de dados, como os códigos *American Standard Code for Information Interchange* (ASCII), e os dados de formatação, como, por exemplo, o tipo de fonte, os recuos e os estilos; e

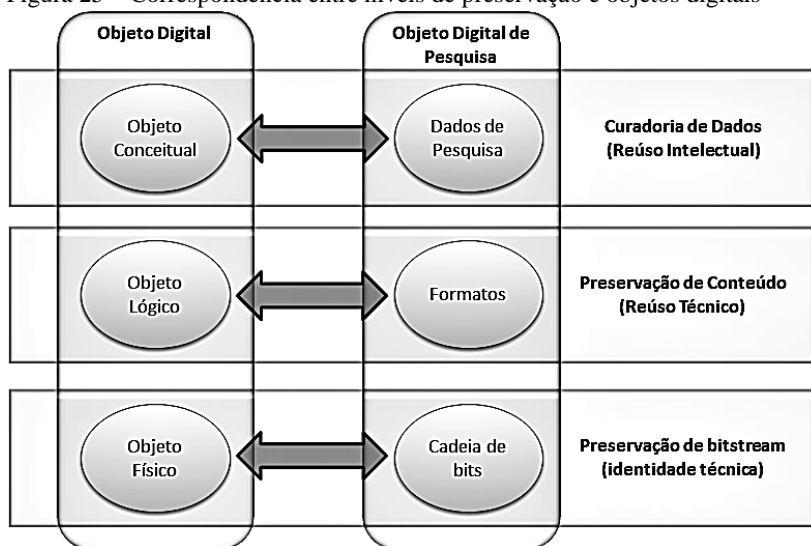
- c) **objeto conceitual** – como objeto conceitual, o objeto digital é reconhecido e entendido por uma pessoa ou, em alguns casos, reconhecido e processado por uma aplicação de computador. É o objeto “do mundo real”, reconhecido como uma unidade significativa de informações, tal como um livro, um contrato, um mapa ou uma fotografia. O conteúdo e a estrutura de um objeto conceitual devem ser contidos de alguma forma no objeto lógico ou nos objetos que representam o objeto na forma digital. No entanto, o mesmo objeto conceitual pode ser representado em diferentes codificações digitais. Por exemplo, um mesmo documento, gravado e reproduzido no processador de texto MS-WordTM e no Adobe Portable Document FormatTM (pdf), pode manter o mesmo conteúdo, aparência e estrutura. Isto é, têm a mesma aparência visual. Este exemplo revela dois importantes aspectos dos objetos digitais e que têm implicações na sua preservação. Primeiro, existem diferentes codificações digitais do mesmo objeto conceitual; e, segundo, que diferentes codificações podem preservar as características essenciais do objeto conceitual.

A National Library of Australia (2003) adota a proposta de Thibodeau (2002), porém, acrescenta uma classe denominada de **elementos essenciais**. Essa classe adicional visa conter a mensagem, o propósito ou as características por que foi decidido preservar o material.

Na preservação digital, essa natureza multicamada dos objetos digitais tem profunda implicação, pois, para cada camada, a preservação significa coisas diferentes (NATIONAL LIBRARY OF AUSTRALIA, 2003).

De acordo com Harvey (2005, p. 46), os conceitos de Thibodeau (2002) são a chave para compreensão da complexidade de preservar o material digital. Harvey (2005, p. 46) denomina o objeto físico de artefato digital e o objeto lógico de objeto digital.

Figura 25 – Correspondência entre níveis de preservação e objetos digitais



Fonte: Baseado em Ludwig (2010).

No projeto WissGrid¹³, Ludwig (2010) identifica três diferentes aspectos da preservação de longo prazo e faz correspondência com as três camadas propostas por Thibodeau (2002):

- a) a preservação de *bitstream* lida com o objeto físico e assegura a estabilidade da cadeia bits, embora o objeto físico possa mudar;
- b) a preservação de conteúdo tenta manter o conteúdo estável, embora os formatos e o processamento técnico possam mudar;
- e
- c) a curadoria de dados permite adições (como, por exemplo, a adição de informações de representação semântica ou informações de contexto) e mudanças no objeto conceitual a fim de mantê-lo utilizável e significativo para uma comunidade.

Segundo Ludwig (2010), o conceito de objeto digital complexo é usado, principalmente, para designar objetos digitais que consistem de

¹³ Projeto financiado pelo Ministério da Educação da Alemanha (www.wissgrid.de), visando ao estabelecimento uma estrutura organizacional e técnica de longo prazo. Entre as principais tarefas, está a preservação de longo prazo de dados de pesquisa.

outros objetos digitais, mas que outros tipos de complexidade podem ser identificados, considerando que:

- o objeto conceitual pode ser parte de outros objetos conceituais;
- o objeto conceitual pode consistir de muitos arquivos;
- os componentes de um objeto digital podem estar distribuídos em diferentes meios, redes ou instituições;
- os objetos podem ser dinâmicos, podendo sofrer mudanças dependendo do tempo e do ambiente;
- os objetos podem ser ativos, como documento e *software* interativo;
- existe a dependência implícita de contexto (técnico, social ou intelectual), como ambiente técnico, direitos/leis e requisito de conhecimento.

Preservar os objetos físico e lógico não garante a obtenção do objeto conceitual. É necessário preservar uma forma de traduzir o objeto lógico para o conceitual.

Nesse sentido, Duranti e Thibodeau (2006) identificam que documentos digitais são codificados com três tipos de dados, em: o dado de conteúdo – que constitui o conteúdo de um documento; o dado da forma – que possibilita ao sistema reproduzir o documento na forma correta; e o dado de composição – que permite determinar o conteúdo e a forma de cada documento.

Como é impossível manter o material na sua forma original, seus atributos devem ser preservados a fim de garantir a autenticidade do material digital (HARVEY, 2005). São as propriedades significativas dos objetos digitais.

2.7.2 Propriedades significativas dos objetos digitais

Propriedades significativas do objeto digital são aquelas que afetam sua qualidade, usabilidade, visualização e comportamento (HEDSTROM; LEE, 2002).

Devido à impossibilidade de manter o material digital na sua forma original, é necessário identificar os atributos do material que são necessários preservar para garantir a autenticidade. Isto tem sido descrito como essência, atividades essenciais, propriedades significativas, elementos essenciais e desempenho (HARVEY, 2005).

O Reino Unido desenvolveu, entre 2007 e 2009, um projeto específico para desenvolver o conceito e uma metodologia para

determinar as propriedades significativas de objetos digitais. Foi financiado pela JISC e conduzido por três instituições: o *Arts and Humanities Data Service* (extinto em 2008), o *King's College London* e o *The National Archives* (TNA). O projeto foi denominado *Investigating Significant Properties of Electronic Content* (InSPECT)¹⁴.

O InSPECT assumiu o termo “propriedades significativas” como preferencial, embora reconheça que os termos características significantes ou essência podem ser usados como sinônimos. Para o propósito do InSPECT propriedades significativas, foi definido como: “as características de objetos digitais que devem ser preservadas ao longo do tempo, a fim de garantir a acessibilidade contínua, a usabilidade e o significado dos objetos” (WILSON, 2007, tradução nossa).

O InSPECT define cinco categorias de propriedades significativas (Quadro 19).

Quadro 19 – Categorias de propriedades significativas

Categoria	Exemplo
Conteúdo	texto, imagem, slides, etc.
Contexto	quem, quando, por quê
Aparência	fonte, tamanho, cor, leiaute, etc.
Estrutura	arquivos incorporados, paginação, cabeçalho, etc.
Comportamento	hiperligações, atualização de cálculos, etc.

Fonte: Baseado em Wilson (2007).

Outros tipos de propriedades, como, por exemplo, as propriedades técnicas e de representação que são significativos para a preservação, não estão dentro do escopo do projeto InSPECT (WILSON, 2007).

Maior detalhamento dos tipos de dados necessários para preservação, seus conceitos e relacionamentos são definidos num modelo que se torna um padrão para a concepção de repositórios digitais: o modelo de referência OAIS.

2.8 MODELO DE REFERÊNCIA OAIS

¹⁴ O material produzido pelo projeto InSPECT está disponível em linha em: <http://www.significantproperties.org.uk/>

O Modelo de Referência *Open Archival Information System* (OAIS) é uma iniciativa para definir um modelo conceitual para a gestão de materiais digitais de valor permanente e estabelece um vocabulário (NATIONAL LIBRARY OF AUSTRALIA, 2003). Esse estabelecimento de uma linguagem comum para a discussão da preservação digital é uma das contribuições do modelo OAIS (HARVEY, 2005; ICPSR, 2009).

O *Consultative Committee for Space Data Systems* (CCSDS), baseado na *National Aeronautics and Space Administration* (NASA), desenvolveu o modelo de referência OAIS como primeiro passo para produção de um padrão formal para o arquivamento de longo prazo de dados de ciência espacial (BALL, 2006; HARVEY, 2005). O modelo de referência tornou-se um padrão internacional, em 2003 (ISO 14721), e, em 2007, foi aprovado como norma brasileira (NBR 15472).

O modelo de referência OAIS é um arcabouço conceitual de alto nível que pode ser usado como um ponto de referência para aqueles que concebem, avaliam ou utilizam sistemas de informação para arquivamento de longo prazo de objetos digitais. Ele é independente da solução de implementação, isto é, não é uma especificação de aplicação e não fornece instruções de como preservar a informação digital (NATIONAL LIBRARY OF AUSTRALIA, 2003).

A principal finalidade do modelo de referência OAIS é o de facilitar um entendimento geral do que é necessário para preservar e acessar informações visando o longo prazo (CCSDS, 2012).

O modelo define que o OAIS é um arquivo, que consiste de uma organização de pessoas e sistemas, com a responsabilidade de preservar informação e torná-la disponível para uma comunidade (CCSDS, 2012). Portanto, quando o texto denomina simplesmente de “OAIS” ou de “um OAIS”, está fazendo menção ao arquivo concebido em conformidade com o modelo. A referência ao arcabouço conceitual é feita explicitamente como “modelo de referência OAIS”.

O termo “Open” é usado para sugerir que a recomendação OAIS e outras recomendações relacionadas sejam desenvolvidas em fóruns abertos, e isso não implica que o acesso ao arquivo seja irrestrito (CCSDS, 2012).

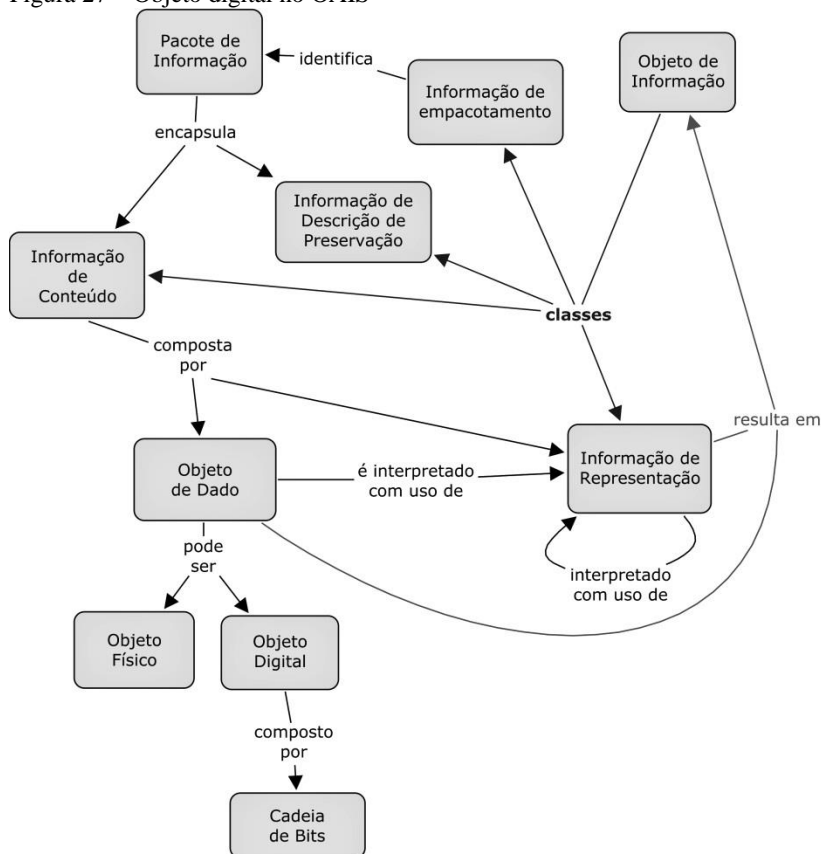
O modelo de referência OAIS identifica entidades funcionais de um sistema de repositórios digitais, suas interfaces internas e externas, e fluxos de informações (Figura 26).

O Modelo de Referência OAIS define objeto digital como: “Objeto composto por um conjunto de cadeias de bits” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2007; CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS - CCSDS, 2012).

Na concepção do OAIS, o objeto digital e a informação de representação (que o interpreta) constituem a informação de conteúdo. Por sua vez, as informações de conteúdo, de descrição de preservação e de empacotamento constituem o pacote de informação. Dessa forma, para o OAIS, o objeto digital que é somente um conjunto de bits que necessita da informação de representação para ser decodificado e interpretado.

O mapa conceitual da Figura 27 explicita a relação existente entre elementos do OAIS e situa o objeto digital no contexto.

Figura 27 – Objeto digital no OAIS



Fonte: Elaborado pelo autor com base no CCSDS (2012)

O conceito que a informação é a combinação de dados e informação de representação é básico no modelo de referência OAIS (PREMIS EDITORIAL COMMITTEE, 2012).

Para o modelo OAIS, as informações de representação são aquelas requeridas para exibição do objeto de dado, ou seja, utilizam-se os conceitos de Thibodeau (2002) para transformar o objeto lógico em objeto conceitual.

O modelo OAIS estabelece que a informação de representação poderá ser substituída por um *software* (CCSDS, 2012) que tenha a capacidade de transformar o objeto de dado em objeto de informação (ou o objeto lógico em conceitual).

As **informações de descrição de preservação**, para o modelo OAIS, são os metadados descritivos, de contexto, de direito de acesso, e outros.

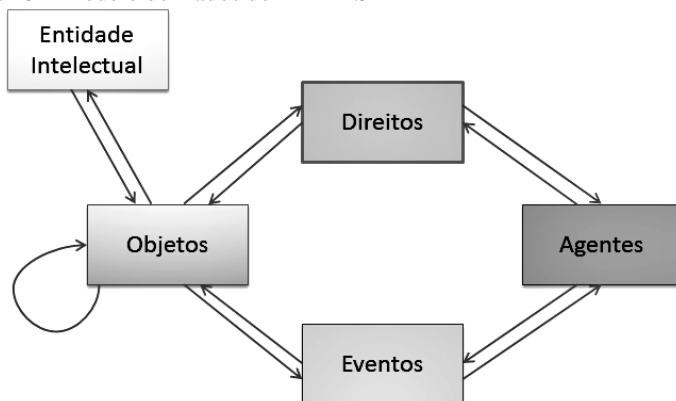
2.9 DICIONÁRIO DE DADOS PREMIS

O *Preservation Metadata: Implementation Strategies* (PREMIS) foi um grupo de trabalho patrocinado pelo *Online Computer Library Center* (OCLC) e *Research Library Group* (RLG), que desenvolveu um conjunto de metadados de preservação, aplicável num amplo leque de contexto de preservação digital, publicado no *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*.

O dicionário de dados PREMIS, atualmente, é mantido pelo PREMIS Editorial Committee do *Library of Congress* e está na versão 2.2 (PREMIS EDITORIAL COMMITTEE, 2012).

O PREMIS teve como foco a implementação, isto é, sempre que possível, as unidades semânticas seriam atendidas por processos automáticos, dispensando a intervenção humana na análise. O grupo de trabalho deu preferência à denominação “unidade semântica” em vez de “elemento de metadados” devido à ênfase na necessidade de saber e não na necessidade de representar (PREMIS EDITORIAL COMMITTEE, 2012). É uma diferença sutil, porém importante. Uma unidade semântica é uma peça de conhecimento, um elemento de metadados é uma maneira definida de representar essa informação num registro de metadados (CAPLAN, 2008).

Figura 28 – Modelo de Dados do PREMIS



Fonte: Adaptado de Premis Editorial Committee (2012).

Um dos princípios mais importantes do dicionário de dados PREMIS é a necessidade de muita clareza no que se descreve. Com esse propósito, define um modelo de dados (Figura 28) para organizar as unidades semânticas no dicionário de dados, com cinco entidades (PREMIS EDITORIAL COMMITTEE, 2012):

- a) Entidade intelectual – um conjunto de conteúdo considerado uma unidade intelectual para o propósito de gestão e descrição, como, por exemplo, livros, fotografia, mapa, base de dados. Uma entidade intelectual pode incluir outra entidade intelectual, por exemplo, uma página web que inclui uma imagem. Uma entidade intelectual pode ter uma ou mais representações digitais;
- b) Objeto ou Objeto Digital – uma unidade discreta de informação na forma digital;
- c) Evento – uma ação que envolve ou afeta um objeto ou agente no repositório de preservação;
- d) Agente – uma pessoa, organização ou *software* associado(a) aos eventos durante a vida de um objeto ou aos direitos associados a ele;
- e) Direitos – são asserções ligadas a um ou vários direitos ou permissões pertencentes a um objeto ou agente.

Segundo o dicionário de dados PREMIS (2012), a necessidade maior especificidade ocasionou diferenças de terminologia em relação ao Modelo de Referência OAIS. Situação considerada corriqueira da transição de um modelo conceitual para uma visão de implementação.

O dicionário de dados PREMIS define objeto digital (ou simplesmente objeto) como uma unidade de informação discreta na forma digital. Um objeto digital pode ser uma representação (*representation*), um arquivo (*file*), uma cadeia de bits (*bitstream*) ou uma cadeia de arquivos (*filestream*). Esta definição é diferente da definição frequentemente utilizada no campo da biblioteca digital, em que o objeto digital é uma combinação de identificador, dados e metadados (PREMIS EDITORIAL COMMITTEE, 2012).

Os quatro tipos de objetos digitais têm o seguinte entendimento no dicionário de dados PREMIS:

- representação – conjunto de arquivos, incluindo metadados estruturais que são necessários para uma representação completa de uma entidade intelectual;
- cadeia de bits – dados de um arquivo que possuem propriedades comuns em relação aos objetivos de

preservação. Um *bitstream* não pode ser transformado em um arquivo independente, sem a adição de dados estruturais;

- arquivo – uma sequência ordenada de bytes reconhecida pelo sistema operacional;
- cadeia de arquivos – é uma cadeia de bits incorporado em um arquivo que pode ser transformado em um arquivo independente, sem o acréscimo de informação adicional, como, por exemplo, uma imagem TIFF incorporada em um arquivo com formato .tar (Tape ARchive).

O PREMIS (2012) categoriza, ainda, os objetos digitais em simples e composto: o primeiro é o objeto de um único arquivo, por exemplo, um documento em um único arquivo PDF; e o segundo, o objeto formado por vários arquivos, por exemplo, uma página web composta por arquivos de imagem e vídeo.


2.10 INTEROPERABILIDADE

Um dos fatores que contribuem para a atual conjuntura de ininterrupta alta produção e consumo de informação é a convergência tecnológica. Um vasto cenário de dispositivos que podem desempenhar funções semelhantes ou complementares é o conceito de convergência tecnológica (MORAIS SQUIRRA, 2005), a exemplo de *smartphones* e televisores que executam aplicativos que, há poucos anos, eram exclusividade de computadores.

A convergência tecnológica exige a capacidade de mobilidade, portabilidade (de aplicações e conteúdos), interconexão e interoperabilidade entre plataformas e operadores (SERRA, 2006).

Na TIC, a interoperabilidade é entendida, predominantemente, como a habilidade de transferir dados e informações entre sistemas (GASSER; PALFREY, 2007; SAEKOW; JIRACHIEFPATTANA, 2011; TAMBOURIS *et al.*, 2008).

Quadro 20 – Níveis de Interoperabilidade Conceitual



Nível	Nome do nível	Premissa (ter em comum)	Definição de informação
6	Conceitual	Modelo conceitual	Pressupostos, restrições, etc
5	Dinâmica	Modelo de execução	Efeito dos dados
4	Pragmática	Modelo de workflow	Uso dos dados
3	Semântica	Modelo de referência	Significado dos dados
2	Sintática	Estrutura de dados	Dados estruturados
1	Técnica	Protocolo de comunicação	Bits e bytes
0	Não	Sem conexão	Não

Fonte: Adaptado de Wang, Tolk e Wang (2009).

O heterogêneo e complexo ambiente imposto pela convergência tecnológica exige melhor entendimento do conceito de interoperabilidade. Neste sentido, Gradmann (2008) identifica quatro dimensões da interoperabilidade, com níveis crescentes de abstração: 1º) técnico; 2º) sintático; 3º) funcional; e 4º) semântico. Na busca por maior compreensão e detalhamento, o *Levels of Conceptual Interoperability Model* (LCIM) propõe sete níveis de interoperabilidade (TOLK; DIALLO; TURNITSA, 2007; TOLK; MUGUIRA, 2003; WANG; TOLK; WANG, 2009).

Para fins da preservação digital, os conceitos e as reflexões acerca da interoperabilidade são também de suma importância. A preservação de longo prazo pode ser vista como uma interoperabilidade ao longo do vetor tempo (GRADMAN, 2008).

2.10.1 Interoperabilidade Temporal

No contexto da preservação digital, a interoperabilidade por meio do estabelecimento de padrões e da migração de formatos de arquivos é problemática, pois o foco está em tornar o recurso preservado compatível com a tecnologia corrente, em oposição à visão de buscar

formas de interoperação de velhas tecnologias com as novas (HEDSTROM, 2001).

A Interoperabilidade Temporal é a habilidade de sistemas atuais ou legados serem interoperáveis com futuros sistemas, que podem fazer uso de novos formatos, modelos de dados, linguagens, protocolos de comunicação e *hardware* (HEDSTROM, 2001), ou seja, é a capacidade de sistemas e coleções de documentos permanecerem acessíveis ao longo do tempo.

Os usuários podem ter a expectativa de que os sistemas não só interoperem entre organizações, nações, línguas e culturas, mas também sejam capazes de interoperar entre o passado e o futuro (HEDSTROM, 2001).

Com a introdução do conceito de interoperabilidade temporal, surge a necessidade de estabelecer mecanismos para que os modelos e padrões de interoperabilidade tenham aplicabilidade contínua, não somente para um determinado período de tempo.

Embora a interoperabilidade temporal seja um conceito em formação (HEDSTROM, 2001), proporciona novas perspectivas à preservação digital, ao examiná-la como um problema de interoperabilidade

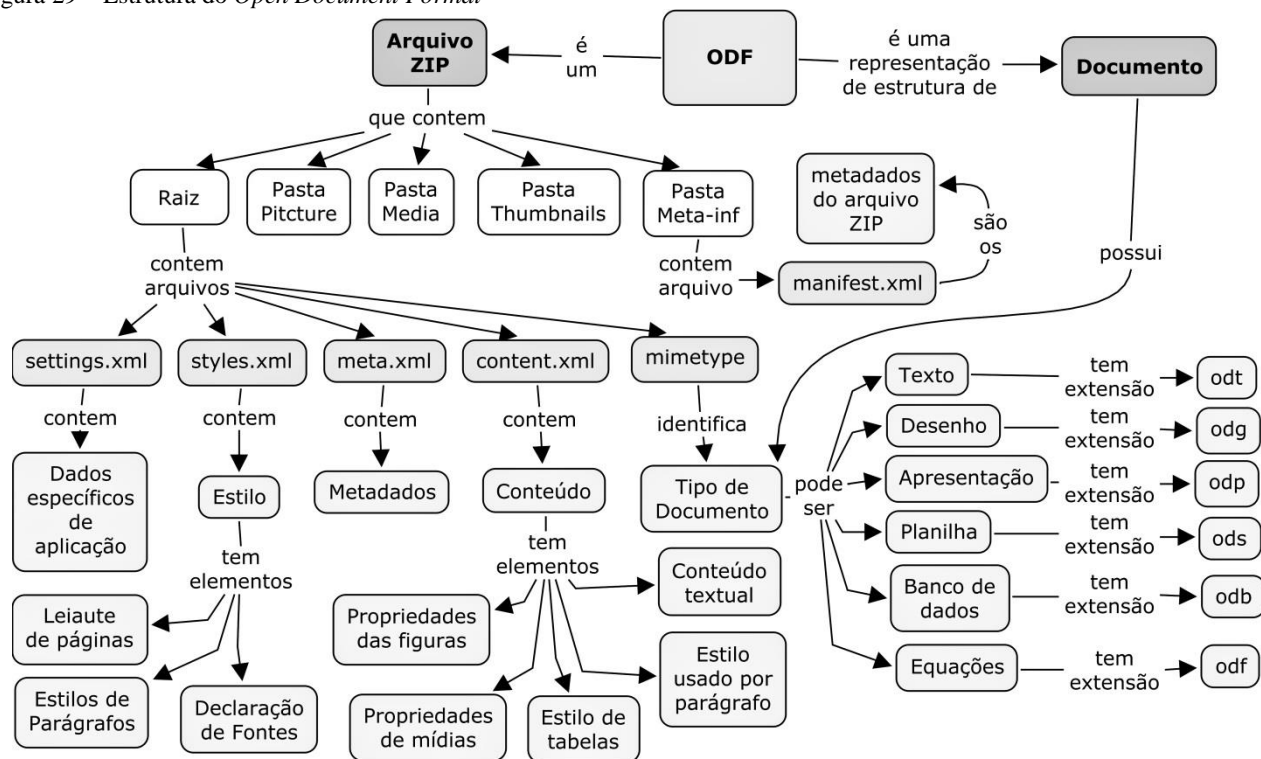
Interoperar com o passado é um desafio complexo, com o futuro, é, aparentemente, um problema insolúvel, especialmente se a abordagem for a manutenção da interoperabilidade com a adoção de uma tecnologia perene. Por outro lado, a preservação da acessibilidade com base na preservação das informações de representação e descrição, juntamente com o conteúdo, pode ser vista como uma forma de interoperabilidade com o futuro.

Neste sentido, padrão de formato de arquivo¹⁵ como o *Open Document Format* (ODF) e o modelo de referência OAIS podem trazer importantes contribuições.

2.11 OPEN DOCUMENT FORMAT – ODF

O ODF é um formato concebido em linguagem XML, desenvolvido pela *Organization for the Advancement of Structured Information Standards* (OASIS), que se tornou padrão ISO/IEC 26300 em 2006.

¹⁵ Um formato de arquivo define a estrutura e os tipos de dados armazenados em um arquivo http://www.techterms.com/definition/file_format

Figura 29 – Estrutura do *Open Document Format*

Fonte: Yamaoka, Gauthier (2013).

O ODF é uma representação idealizada de estrutura de documento. Um arquivo ODF é internamente composto de vários arquivos encapsulados em um Java Archive (JAR), que é um arquivo compactado no formato ZIP, portanto, não é totalmente correto chamar de “arquivo” Open Document (EISENBERG, 2005), denominação que foi adotada para fins de simplificação.

A Figura 29 é um mapa conceitual que apresenta a estrutura e os principais elementos do arquivo ODF.

A estrutura representada na Figura 29 é facilmente visualizada descompactando um arquivo ODF, a exemplo da Figura 30 que apresenta a pasta correspondente a um arquivo de texto no formato ODF (com extensão .odt) descompactado.

Figura 30 – Exemplo de um arquivo ODF descompactado.

Nome	Tamanho	Tipo	Data de modificação
Configurations2	0 byte	Pasta	
Media	38,3 MB	Pasta	
META-INF	1,4 KB	Pasta	
Pictures	147,5 KB	Pasta	
Thumbnails	18,8 KB	Pasta	
content.xml	11,7 KB	Documento ...	14 maio 2013, 07:53
manifest.rdf	899 bytes	Arquivo RDF	14 maio 2013, 07:53
meta.xml	1,7 KB	Documento ...	14 maio 2013, 07:53
mimetype	39 bytes	desconhecido	14 maio 2013, 07:53
settings.xml	9,6 KB	Documento ...	14 maio 2013, 07:53
styles.xml	15,7 KB	Documento ...	14 maio 2013, 07:53

Fonte: Yamaoka, Gauthier (2013).

Os principais elementos da estrutura de um arquivo ODF são:

- A pasta-raiz que conta com os seguintes arquivos:
 - a. settings.xml – com os dados específicos da aplicação (por exemplo, do LibreOffice) que gerou o documento;
 - b. styles.xml – que define os estilos utilizados no documento, isto é, são os elementos que definem a formatação do documento, como o leiaute das páginas, estilos dos parágrafos e as declarações de fontes utilizadas;

- c. meta.xml – contém os metadados relacionados ao documento, como datas, criador, assunto, palavras-chave, etc.
- d. content.xml – é o conteúdo do documento;
- e. mimetype – especifica o MIME Type do documento (por exemplo, application/vnd.oasis.opendocument.text para documentos texto que são identificados com a extensão .odt no ODF).
- A pasta Picture contém os arquivos de imagens utilizadas no documento;
- A pasta Media contém os arquivos de mídia como vídeos e áudios inseridos no documento;
- A pasta Meta-Inf contém o arquivo manifest.xml que lista os arquivos contidos no arquivo ODF. São os metadados do pacote zipado;
- A pasta Thumbnails contém uma imagem miniatura do documento.

2.11.1 Representação e descrição no ODF e OAIS

Um documento XML – como o ODF – tem como característica ser inteligível a humanos, independente da disponibilidade do *software* para exibição do documento, isto é, mesmo sem o *software* para exibição, as pessoas terão acesso ao conteúdo, embora sem a formatação.

O ODF tem também como característica a estrutura do documento explícita e interpretável. É perceptível que a estrutura do documento ODF possui semelhanças com o modelo de informação estabelecido no modelo de referência OAIS, como ilustrado na Figura 31 e no

A relação entre os componentes da modelo de referência OAIS e um arquivo ODF é:

Quadro 21 – Correspondência entre a estrutura ODF e OAIS

Componente do Modelo de Referência OAIS	Elementos arquivo ODF
Pacote de Informação	Arquivo ODF (arquivo zip)
Descrição do Pacote	manifest.xml
Informação de Empacotamento	
Informação de Descrição de Preservação	meta.xml
	Pasta thumbnails
Informação de Representação	settings.xml
	styles.xml
	mimetype
Objeto de Dado	conteudo.xml
	pasta media
	pasta picture

Fonte: Elaborado pelo autor.

A correspondência da estrutura do documento ODF com o modelo de informação preconizado pelo modelo OAIS pode trazer importantes benefícios na busca pela interoperabilidade temporal dos objetos digitais.

O modelo OAIS estabelece uma terminologia comum, já disseminado na comunidade interessada em preservação digital. As conformidades entre os componentes dos modelos podem ser facilmente identificadas, o que facilita o processo de interpretação do conteúdo e informações de descrição e representação do objeto digital.

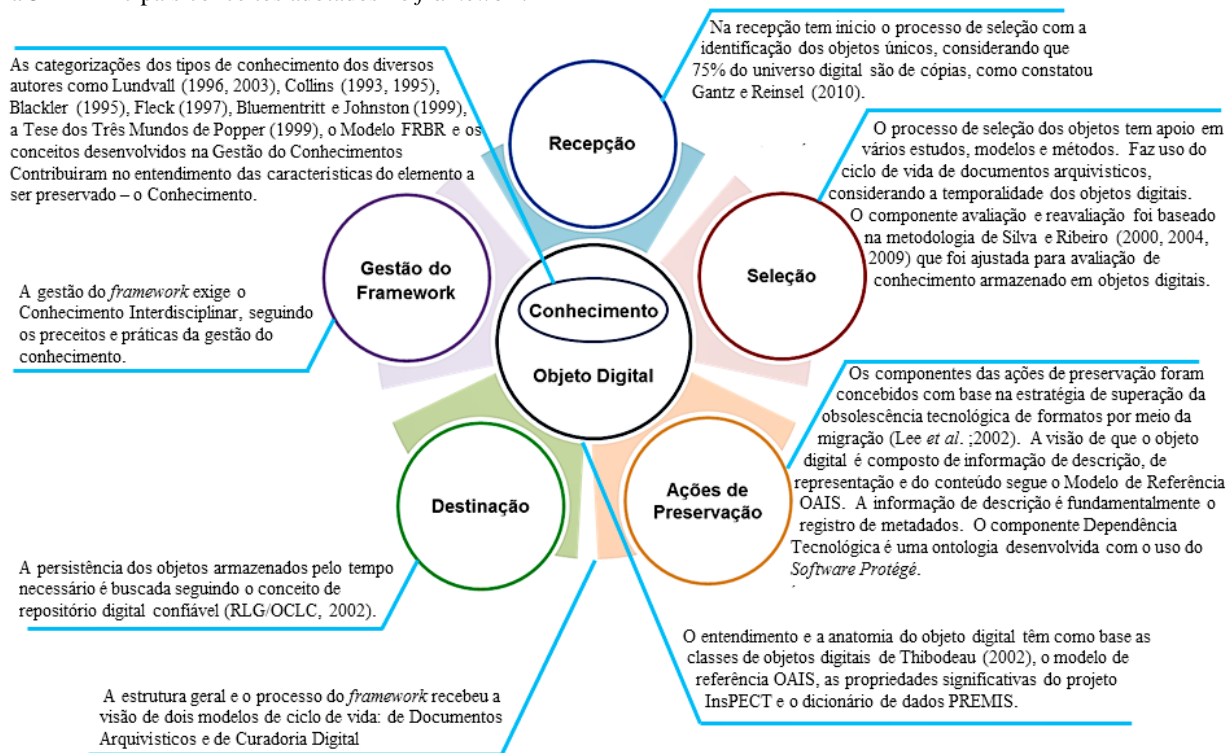
Mesmo que, no futuro, ambos os modelos venham a sofrer alterações ou, até mesmo, sejam descontinuados, o fato de ambos serem padrões ISO permitirá a fácil recuperação das documentações de especificação. Adicionalmente, como o arquivo ODF permite incluir novos elementos de metadados no arquivo meta.xml, a conformidade do ODF com o modelo de referência OAIS poderá ser ainda maior, contribuindo para a interoperabilidade temporal e a preservação do objeto digital criado nesse formato.

3 METODOLOGIA E ALINHAMENTO TEÓRICO

A elaboração do *framework* seguiu as etapas apontadas na seção 1.

Na primeira etapa a pesquisa bibliográfica permitiu o aprofundamento de teorias, técnicas, metodologias e modelos aplicáveis no *framework*.

O infográfico (Figura 32) ilustra os principais conceitos adotados na concepção do *framework*. Apresenta no seu centro o elemento de preservação – o Objeto Digital com Conhecimento e os cinco grupos de componentes: Recepção, Seleção, Ações de Preservação, Destinação e Gestão do *Framework*.

Figura 32 – Principais conceitos adotados no *framework*

Fonte: Elaborado pelo autor

Na segunda etapa, foram identificados e concebidos dois componentes centrais do *framework*: o componente avaliação e reavaliação e o componente dependência tecnológica.

3.1 AVALIAÇÃO E REAVALIAÇÃO

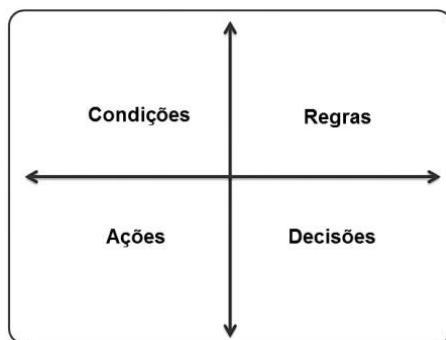
A base do componente avaliação e reavaliação é a metodologia de metodologia de Silva e Ribeiro (2000, 2004, 2009) que foi analisada e ajustada para uso no *framework*.

Inicialmente, com a finalidade de visualizar e analisar os critérios de avaliação de Silva e Ribeiro (2000, 2004, 2009) - densidade, pertinência e frequência - foram plotados em um gráfico tridimensional X, Y e Z.

Com vistas à desejada objetividade e à possibilidade de implementações automatizadas do processo de avaliação, os parâmetros e critérios da operação metodológica foram representados em uma tabela de decisão. A tabela de decisão é um instrumento que permite verificar todas as combinações possíveis em um conjunto de condições e seus resultados (BEATTY; CHEN, 2012, p. 234).

Uma tabela de decisão possui quatro quadrantes, com as condições, as ações, as regras e os resultados, dispostos conforme Figura 33.

Figura 33 – Quadrantes de Tabelas de Decisão



Fonte: Elaborado pelo autor.

3.2 ONTOLOGIA DE DEPENDÊNCIA TECNOLÓGICA

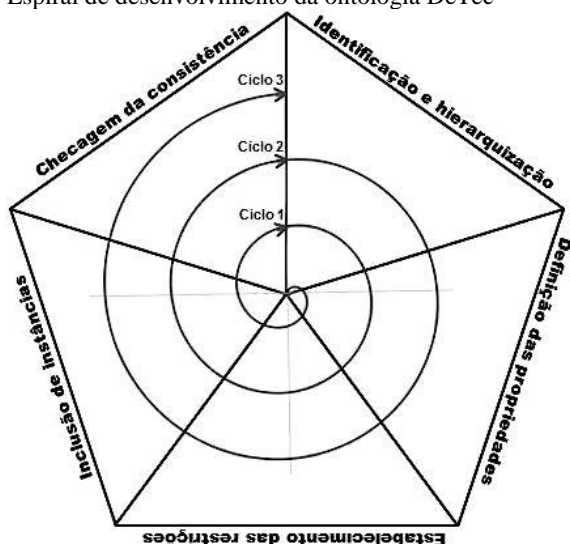
A Ontologia de Dependência Tecnológica (DeTec) foi modelada com o uso do Protégé¹⁶ na versão Protégé-OWL 3.4.8.

Para a construção da DeTec, foram seguidos os passos tradicionalmente utilizados para construção de ontologias. O domínio a ser modelado já estava definido, portanto, o processo seguido foi:

- a) identificação das classes e a hierarquização;
- b) identificação e definição das propriedades das classes;
- c) estabelecimento das restrições;
- d) inclusão de instâncias; e
- e) checagem da consistência da ontologia.

A ontologia DeTec foi concebida no modelo em espiral, isto é, em ciclos. Em cada ciclo, foram executados os cinco passos do processo, com correções e ajustes para aperfeiçoamento da ontologia (Figura 34).

Figura 34 – Espiral de desenvolvimento da ontologia DeTec



Fonte: Elaborado pelo autor

Os demais componentes do *framework* e seus inter-relacionamentos foram concebidos na terceira etapa.

¹⁶ <http://protege.stanford.edu/>

Na quarta etapa a coerência e aplicabilidade do *framework* foram verificadas por meio de entrevistas realizadas com especialistas.

3.3 VERIFICAÇÃO E AJUSTES DO *FRAMEWORK*

A verificação do *framework* foi realizada a partir do conhecimento de especialistas que atuaram como agentes avaliadores da proposta, portanto, uma pesquisa qualitativa, tendo em vista que a análise e conclusões deste trabalho tiveram como base o parecer dos especialistas convidados, detalhados nas subseções subsequentes.

Para coletar as informações, foi utilizada a técnica da entrevista semiaberta. A entrevista tem a vantagem de oferecer maior amplitude por não estar presa a um documento, como no questionário (LAVILLE; DIONNE, 1999, p.187).

A entrevista semiaberta é utilizada para pesquisa qualitativa, com questões semiestruturadas, com o uso de um roteiro-base e com abordagem "em profundidade" (DUARTE, 2006, p. 65).

Para garantir a uniformidade do processo, foram preparados *slides* com explicação do *framework* que foram apresentados aos entrevistados, previamente à aplicação da entrevista.

A apresentação seguiu a seguinte estrutura:

- Problema de pesquisa;
- Objetivo do projeto;
- Conceito de objeto digital;
- Curadoria e preservação digital;
- Componentes e processo do *Framework*.

O roteiro e as questões da entrevista encontram-se no Apêndice A.

Como treinamento e aprimoramento do instrumento de pesquisa, foi realizado um pré-teste com dois especialistas: um com perfil de especialista em computação e outro com perfil de especialista em arquivística, com conhecimento em preservação de documentos arquivísticos digitais.

No pré-teste, foram avaliados: a clareza das perguntas, o tempo de aplicação da entrevista, a abrangência das respostas obtidas e o processo de gravação da entrevista.

A entrevista semiaberta, geralmente, conta com poucas questões (de quatro a sete), mas suficientemente amplas para serem discutidas em profundidade, sem interferência entre elas (DUARTE, 2006, p.66).

A estrutura das entrevistas buscou investigar os seguintes aspectos do *framework*:

- Aplicabilidade do *framework*;
- Eficácia do processo de seleção de objetos digitais;
- Importância das ações de preservação;
- Alinhamento do *framework* em relação ao objetivo proposto;
- Críticas e sugestões.

No estudo qualitativo, a qualidade antepõe a quantidade. Duarte (2006, p. 68) recomenda que "[...] são preferíveis poucas fontes, mas de qualidade, a muitas, sem relevo. Desse modo, e no limite, uma única entrevista pode ser mais adequada para esclarecer determinada questão do que o censo nacional."

Devido às características do problema que é emergente, de alta complexidade e o caráter interdisciplinar do assunto, foram considerados três perfis de especialista para a entrevista: especialista em tecnologia da informação, especialista em Gestão do Conhecimento e especialista em preservação de documentos arquivísticos digitais.

Os requisitos comuns para os três perfis foram:

- formação: preferencialmente com pós-graduação; e
- experiência: ter participado ativamente em atividades como: desenvolvimento de projetos, pesquisas e grupos de trabalhos.

A lista inicial contou com uma lista com vinte e sete especialistas como candidatos a serem entrevistados (dez especialistas em tecnologia da informação, dez especialistas em gestão do conhecimento, e sete especialistas em gestão e preservação de documentos arquivísticos digitais).

Foram selecionados seis especialistas com base nos requisitos e na disponibilidade destes, sendo dois especialistas em tecnologia da informação, dois especialistas em gestão do conhecimento e dois especialistas em gestão e preservação de documentos arquivísticos digitais. O Apêndice B relaciona os perfis dos entrevistados.

As entrevistas foram realizadas entre 23 de setembro e 18 de outubro de 2013. Cinco entrevistas foram realizadas presencialmente e uma entrevista realizada por meio de recursos de webconferência.

Todas as entrevistas foram gravadas para posterior transcrição da síntese (Anexo C – Síntese das respostas) e foram concedidas no próprio local de trabalho dos entrevistados.

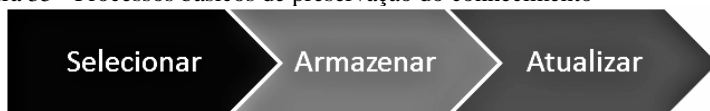
A duração total de cada entrevista foi em torno de 90 minutos, considerando o tempo de apresentação do *framework*, de 30 a 50 minutos, e o tempo da entrevista com duração de 30 a 50 minutos.

Na quinta e última etapa, o *framework* foi ajustado baseado nos pareceres dos especialistas.

4 FRAMEWORK DE PRESERVAÇÃO DE OBJETOS DIGITAIS

Para a preservação do conhecimento, as organizações devem dominar três processos básicos (Figura 35): a seleção; o armazenamento e a atualização da memória organizacional (ROMHARDT, 1997).

Figura 35 - Processos básicos de preservação do conhecimento



Fonte: Adaptado de Romhardt (1997).

Esses três processos básicos estão presentes no *framework* de preservação de objetos digitais proposto neste trabalho, concebido a partir dos fundamentos teóricos apresentados na seção 2.

4.1 DESCRIÇÃO DO FRAMEWORK

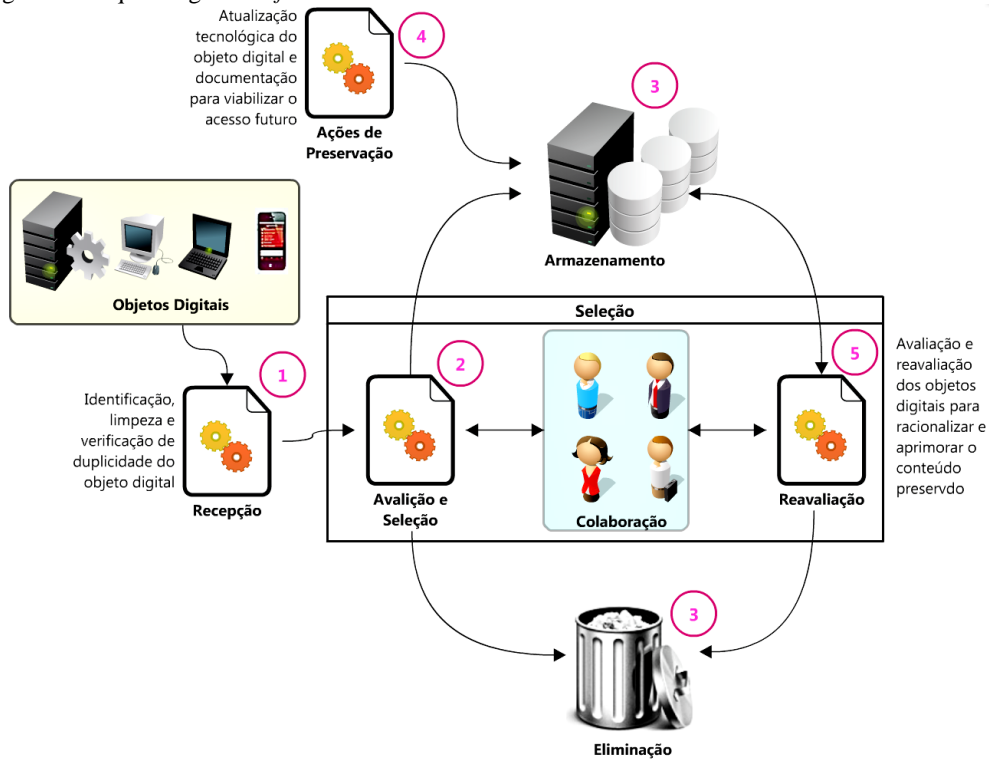
A visão geral do *framework* é apresentada no infográfico (Figura 36) e o diagrama com os componentes (Figura 37).

O infográfico (Figura 36) introduz o *framework* apresentando uma visão macro do processo de seleção e preservação do objeto digital:

- 1) recepção dos objetos digitais;
- 2) o objeto é encaminhado para seleção, realizada com processo colaborativo;
- 3) os objetos selecionados para guarda são encaminhados para o repositório e os considerados desnecessários, para o descarte;
- 4) os objetos do repositório são monitorados quanto às tecnologias necessárias para seu uso;
- 5) os objetos armazenados são, periodicamente, reavaliados quanto à pertinência da sua guarda.

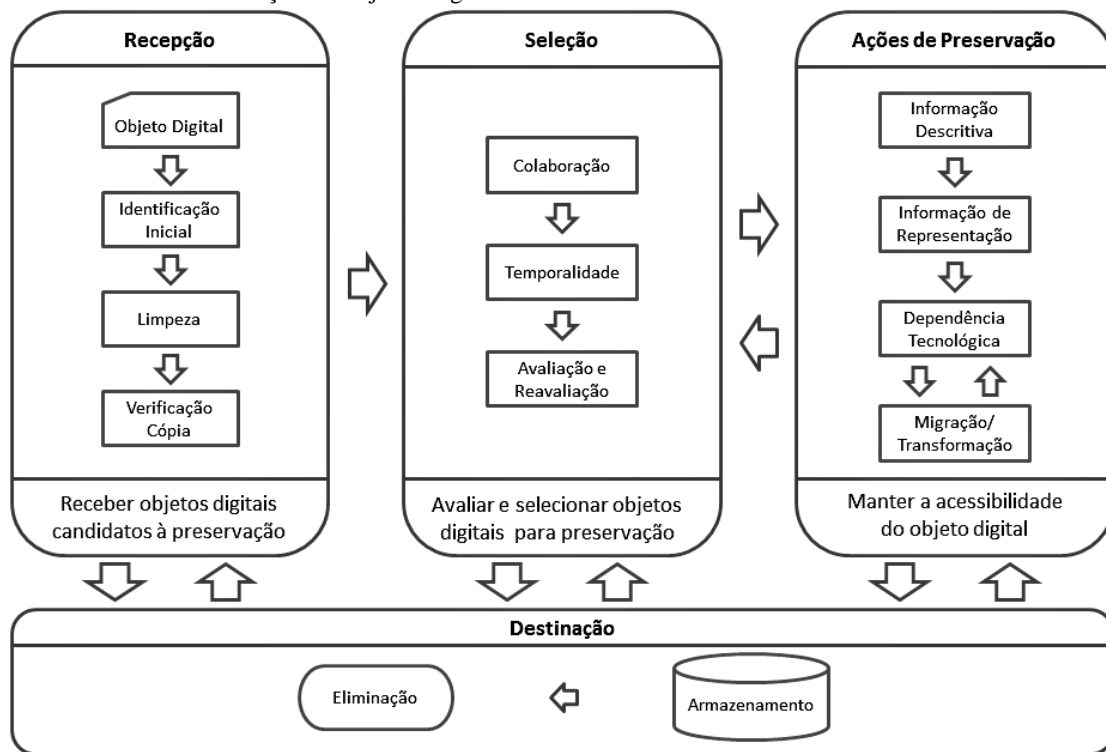
O segundo diagrama (Figura 37) apresenta, ainda, uma visão geral do *framework*, destacando os principais componentes.

Figura 36 – Infográfico do quadro geral do *framework*.



Fonte: Elaborado pelo autor.

Figura 37 – Framework de Preservação de Objetos Digitais



Fonte: Elaborado pelo autor.

O *framework* conta com quatro principais grupos de componentes, que são: Recepção; Seleção; Ações de Preservação; e Destinação.

A Recepção é um grupo de componentes simples, mas indispensável para garantir a segurança e economia dos processos subsequentes. Visa identificar o objeto digital com a extração de metadados, eliminar eventuais códigos maliciosos e reduzir a entrada de cópias de objetos digitais.

A Seleção busca racionalizar o ambiente de armazenamento e aprimorar o conteúdo, selecionando os objetos relevantes para a preservação do conhecimento e descartando aqueles considerados desnecessários, por meio do processo de avaliação e reavaliação com o uso de processo colaborativo.

A Destinação visa operacionalizar as ações decorrentes dos processos de Seleção, que é a eliminação ou o armazenamento do objeto digital.

As Ações de Preservação contam com os processos de documentar o objeto digital com as informações de descrição e de representação, registrar as informações de dependência tecnológica para apoiar o seu processo de gestão e ações de migração dos objetos digitais para novas tecnologias visando superar a obsolescência tecnológica.

Os pilares do *framework* proposto são os grupos "Seleção" e "Ações de Preservação", que também podem ser destacados como contribuição principal que este trabalho agrega no estudo da curadoria e preservação digital.

4.2 COMPONENTES DA RECEPÇÃO

4.2.1 Objeto Digital

A partir dos fundamentos apresentados na seção 2.7, este trabalho assume como características e conceitos de objeto digital:

- a) objeto digital é um objeto de informação, de qualquer tipo e formato, expressa sob a forma digital, que podem ser objeto físico, objeto lógico ou objeto conceitual¹⁷;
- b) objeto digital consiste da combinação de dados de conteúdo mais a informação de representação e descrito por informação de descrição¹⁸;

¹⁷ Baseada nos conceitos de Thibodeau (2002).

- c) objeto digital engloba qualquer informação digital binária, que pode ser simples, complexo, *software*, banco de dados;
- d) as propriedades significativas dos objetos digitais – conteúdo, contexto, aparência, estrutura e comportamento – devem ser preservados no objeto original e nos objetos decorrentes da migração do formato original para outros formatos.

A instituição poderá enviar um objeto digital criado, recebido ou adquirido para o *framework* em qualquer momento do seu ciclo de vida, a partir do qual o objeto passará a ser mantido por esse.

4.2.2 Componentes identificação inicial e limpeza

O componente identificação inicial visa coletar os metadados básicos do objeto digital, como o nome do objeto no sistema de arquivos, tamanho, formato, data de criação, data de modificação, etc.

O componente Limpeza tem como objetivo a verificação e limpeza de códigos maliciosos eventualmente presentes no objeto digital.

4.2.3 Componente verificação de cópia

A verificação de cópia é o primeiro passo do processo de seleção de objetos digitais. Visa identificar se o objeto digital é uma cópia idêntica do objeto armazenado em um repositório.

Com esta ação, o *framework* pretende reduzir significativamente esforços das etapas subsequentes, considerando os estudos que 75% do universo digital é formado por cópias de objetos digitais (GANTZ; REINSEL, 2010).

A cópia poderá ser identificada com a extração e verificação de metadados, como nome do arquivo, data de modificação, formato do arquivo, tipo de conteúdo (content-type). O uso de técnicas a soma de verificação (*checksum*) do arquivo proporciona maior exatidão na identificação de duplicatas de arquivos.

Nos casos de alguma dúvida, a verificação final deverá ser feita por humanos.

Cópia idêntica, neste caso, significa que existe um objeto digital lógico idêntico armazenado no repositório, ou seja, objeto digital com o

¹⁸ Baseada nos conceitos do modelo de referência OAIS, porém, neste trabalho, foi adotado o termo "objeto digital", o que o modelo OAIS denomina de "objeto de informação".

mesmo conteúdo e formato. No entanto, se um objeto digital for criado, por exemplo, no formato doc e gravado também no formato pdf, ou uma fotografia no formato tiff e no formato jpg, terão dois objetos lógicos para o mesmo objeto digital conceitual. Nestes casos, a comparação de nome do objeto, tipo de conteúdo e de *checksum* não possibilitarão a identificação que os objetos são, na realidade, cópias. Nestes casos, somente a existência de metadados descritivos corretos ou o desenvolvimento ou aquisição de *softwares* que identificam padrões visuais poderão evitar ou reduzir a verificação humana.

As cópias são encaminhadas para eliminação e o objeto identificado como único para etapa subsequente, que é a gestão da dependência tecnológica (seção 4.4.1)

4.3 COMPONENTES DA SELEÇÃO

Quando se trata do que deve ser preservado ou descartado, observa-se que a dificuldade é semelhante, seja para o arquivista seja para o gestor do conhecimento, portanto, o componente seleção é, provavelmente, o mais complexo para ser operacionalizado.

No processo de preservação do conhecimento, o problema está na seleção entre o conhecimento com tempo de validade limitado e o conhecimento que é absolutamente necessário para o futuro (ROMHARDT, 1997).

O descarte é um ato, geralmente, mais simples em objetos materiais, mas muito complexo quando se trata de conhecimento.

A importância do desaprender é discutida na teoria administrativa, numa tensão entre a destruição e a proteção de velhos conhecimentos (ROMHARDT, 1997).

Desaprender é o processo em que indivíduos e organizações admitem liberar o aprendizado anterior com o propósito de acomodar novas informações e comportamentos (BECKER, 2005). Programar a “desaprendizagem” intencional e consciente, porém, implica na clara identificação do conhecimento obsoleto que pode ser descartado.

Na Arquivologia, há décadas, pesquisadores da área discutem como viabilizar a avaliação e seleção de documentos arquivísticos, surgindo a corrente que a defende e a corrente que a questiona devido a sua complexidade.

Couture (2005) defende que a avaliação é a central e nobre função da prática arquivística contemporânea e, conseqüentemente, o ato de julgar o valor do documento arquivístico é um grande desafio para os profissionais da área.

Por outro lado, o fato da avaliação exigir análises complexas, com possibilidades de erros e enganos por parte dos avaliadores, leva ao raciocínio que, talvez, a preocupação deveria ser na preservação da totalidade (MACEVICIUTE, 2012).

Para fins do *framework* proposto, embora a **Seleção** seja um dos pilares, caso uma instituição opte por preservar a totalidade dos objetos digitais, é exequível a supressão parcial ou total dos componentes da seleção, por exemplo, manter somente a verificação de cópias.

O assunto avaliação suscita ainda duas questões que determinam o raciocínio e o resultado do processo.

A primeira questão é relativa ao objetivo da avaliação, isto é, se a avaliação tem como meta a eliminação ou a preservação. Atinente a isso, Couture (2005) argumenta que a avaliação para identificar documentos para a eliminação e a avaliação para identificar documentos para a preservação são funções diametralmente opostas. Enquanto a primeira escola preocupa-se com interesses de curto prazo, como economia de espaço, interesses administrativos e financeiros, a segunda está preocupada com interesse de longo prazo, o patrimônio e o arquivo como evidência.

O *framework* proposto tem como foco a preservação de objetos digitais visando à preservação do conhecimento, portanto, acompanha, prioritariamente, a meta da preservação, no entanto, sem perder a visão que é necessário descartar os objetos desnecessários visando à sustentabilidade do processo de preservação dos objetos considerados importantes.

A segunda questão refere-se à subjetividade do julgamento. Se é difícil, quiçá impossível o julgamento ser completamente objetivo, a meta é a objetividade relativa, isto é, buscar ser o mais objetivo possível (COUTURE, 2005).

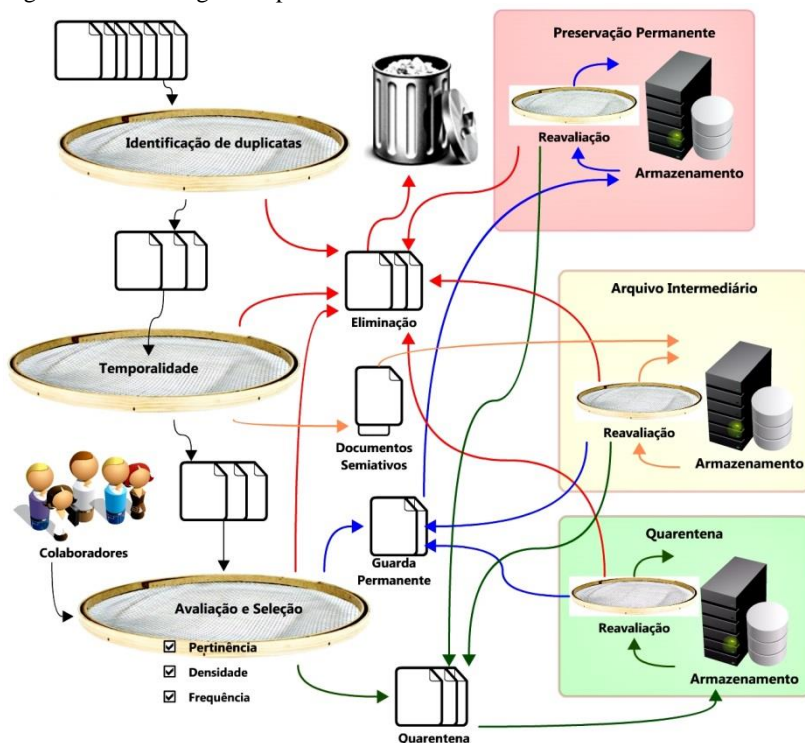
Com o propósito dessa maior objetividade possível no processo de avaliação e seleção, o *framework* estabelece a estratégia de seleção gradativa, ou seja, de "peneiramento sucessivo" (Figura 38).

O primeiro peneiramento é realizado na recepção do objeto digital, que é a verificação se não é cópia de um objeto existente.

Os dois peneiramentos subsequentes visam:

- 1ª) verificar se o objeto já possui temporalidade definida;
- 2ª) realizar a avaliação e seleção.

Figura 38 – Estratégia do "peneiramento sucessivo"



Fonte: Elaborado pelo autor.

Os peneiramentos para o contínuo aprimoramento e atualização da qualidade do conteúdo são realizados nos três tipos de repositórios: de preservação permanente, de arquivo intermediário e de quarentena.

4.3.1 Componente Colaboração

O uso de uma camada de descrição dos objetos digitais, isto é, dos metadados, é considerado a solução para diferentes desafios no contexto da sociedade digital.

A busca e o aprimoramento da encontrabilidade (*findability*), o desenvolvimento da web semântica e a preservação digital são fortemente baseados na utilização de metadados. No entanto, os metadados que podem ser gerados automaticamente por máquinas ainda estão praticamente restritos aos elementos, como formato do arquivo, data de criação, data de modificação e tamanho do arquivo.

Consequentemente, significativa parte do trabalho de criação dos metadados recai para ser executado por humanos, dificultando a operacionalização, pelo alto custo e dificuldade de manter uma equipe dedicada a essa função.

Tendo em vista essas barreiras, o *framework* propõe o uso intensivo de tecnologias de *social computing*.

O *social computing* refere-se aos relacionamentos e à estrutura de poder resultante do uso de tecnologias Web 2.0, como as redes sociais, blogs, microblogs, wikis e tagueamentos (O'DELL; HUBERT, 2011, p.81), que tem como exemplo a conhecida Wikipedia.

O *social computing* permite que os colaboradores participem de acordo com os interesses, conhecimentos e disponibilidades.

A Gestão do Conhecimento está cada vez mais focada em conectar os colaboradores da organização, o que permite a aproximação de pessoas com interesse comum e o acesso seletivo do conteúdo (O'DELL; HUBERT, 2011, p.83).

4.3.2 Componente Temporalidade

O segundo peneiramento no processo de seleção é a verificação da temporalidade em objetos identificados como documentos arquivísticos.

No ciclo de vida documental arquivístico, a avaliação define quais documentos serão preservados para fins administrativos ou de pesquisa e tem como instrumento resultante a tabela de temporalidade, com objetivo de garantir o acesso à informação (ARQUIVO NACIONAL, 2001).

Para os documentos arquivísticos, a tabela de temporalidade e destinação é um instrumento criado e trabalhado em arquivos, que estabelece o prazo que um documento é mantido no seu ciclo de vida, nas fases corrente, intermediária e sua destinação após esse prazo, isto é, para a eliminação ou guarda permanente.

A tabela de classificação, temporalidade e destinação de documentos de arquivo relativos às atividades-meio da administração pública é estabelecida pela Resolução nº 14, de 24 de outubro de 2001, pelo Conselho Nacional de Arquivos (ARQUIVO NACIONAL, 2001).

A tabela das atividades-meio é composta de duas classes: 000 – Administração Geral; e 900 – Assuntos Diversos.

A classe 000 - Administração Geral é dividida em oito subclasses:

- 010 Organização e Funcionamento;

- 020 Pessoal;
- 030 Material;
- 040 Patrimônio;
- 050 Orçamento e Finanças;
- 060 Documentação e Informação;
- 070 Comunicações; e
- 090 Outros assuntos referentes à administração geral.

A classe 900 – Assuntos Diversos, tem como subclasses:

- 910 Solenidades. Comemorações. Homenagens;
- 920 Congressos. Conferências. Seminários. Simpósios. Encontros. Convenções. Ciclos de Palestras. Mesas Redondas;
- 930 Feiras. Salões. Exposições. Mostras. Concursos. Festas;
- 940 Visitas e Visitantes;
- 990 Assuntos Transitórios.

A essas subclasses, a tabela institui 270 assuntos subordinados.

A tabulação e análise da tabela possibilitou concluir, quanto a destinação dos documentos, que:

- para 75,6% dos assuntos, a destinação é a eliminação;
- para 24,4% dos assuntos, a destinação é a preservação permanente;
- as classes com assuntos que têm maior percentual de eliminação são: 070 Comunicações (93,33%), 030 Materiais (91,30%) e 020 Pessoal (85,57%);
- 85,56% dos assuntos requerem, no máximo, 10 anos de tempo de guarda.

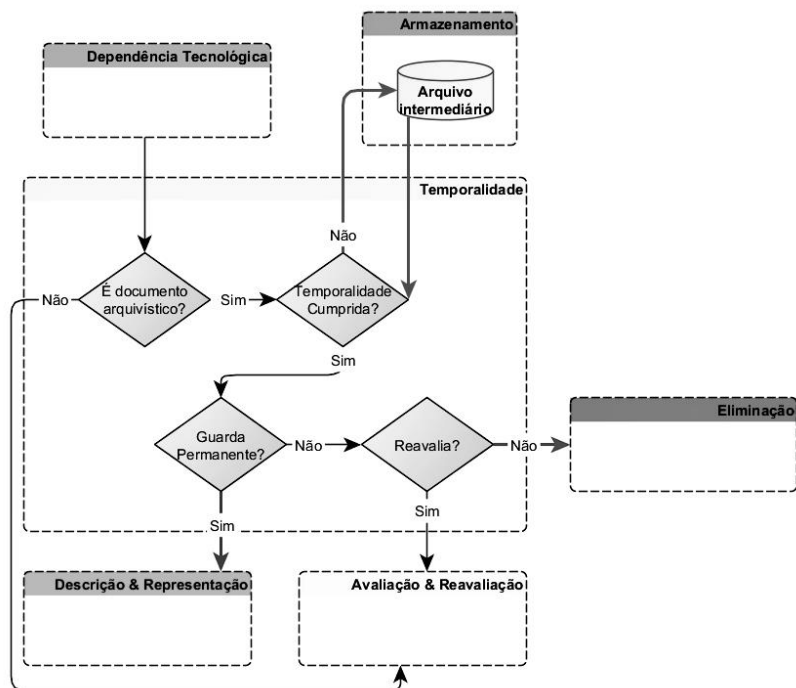
O Conselho Nacional de Arquivos publicou, também, a tabela de temporalidade e destinação de documentos de arquivo relativos às atividades-fim das instituições federais de ensino superior – IFES (ARQUIVO NACIONAL, 2011).

Na tabela destinada às IFES, dos 336 assuntos, 55,36% estão destinados à guarda permanente, em contraste à tabela de atividades-meio que destina somente 24,4% dos assuntos à guarda permanente, confirmando a lógica que os documentos das atividades-fim têm maior valor para a instituição.

Para os objetos digitais considerados documentos arquivísticos, a verificação da temporalidade precede o processo de avaliação e seleção, ou seja, na lógica do *framework*, um objeto digital considerado documento arquivístico não poderá ser eliminado se a tabela de

temporalidade o considera de guarda permanente. Por outro lado, um objeto digital considerado documento arquivístico poderá ser preservado, mesmo que a tabela de temporalidade o considere destinado à eliminação. Essa situação poderá ocorrer se o objeto em questão for considerado importante para fins da preservação do conhecimento organizacional.

Figura 39 – Componente Temporalidade



Fonte: Elaborado pelo autor.

A Figura 39 apresenta o fluxograma do componente Temporalidade, no qual fica explícito que um documento arquivístico destinado à eliminação poderá ser reavaliado para verificar o seu valor no contexto da preservação do conhecimento.

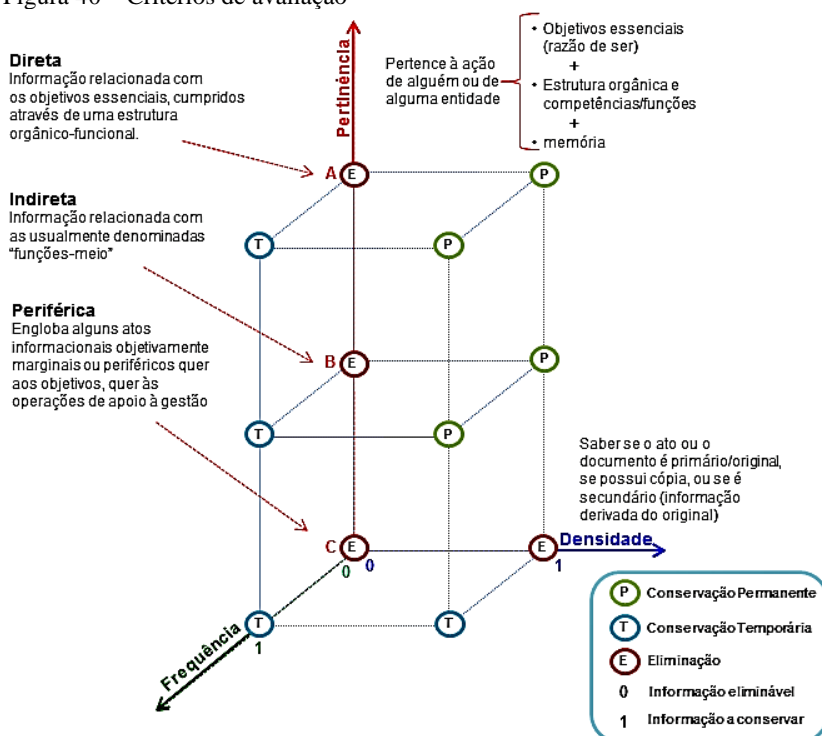
4.3.3 Componente Avaliação e Reavaliação

A Figura 40 elaborada com base no Quadro 15 (página 88), verifica-se no gráfico tridimensional X, Y e Z (respectivamente

densidade, pertinência e frequência), que na metodologia de Silva e Ribeiro (2000, 2004, 2009):

- 1) quando a pertinência for direta ou indireta, a densidade determina a conservação permanente do documento, independente da sua frequência de uso;
- 2) quando a pertinência for periférica, a frequência determina a conservação temporária do documento;
- 3) em qualquer grau de pertinência, se a densidade for zero, a frequência de uso do documento determina a sua conservação temporária.

Figura 40 – Critérios de avaliação



Fonte: Elaborado pelo autor com base em Silva e Ribeiro (2000, 2004, 2009).

Com a aplicação da operação metodológica de avaliação, a tabela de decisão resultante é mostrada no Quadro 22. Para construir a tabela de decisão, o critério pertinência foi alterado em relação à metodologia original para dois valores: 1) tem pertinência; e 2) não tem pertinência.

O primeiro engloba os níveis Direta e Indireta, e o segundo corresponde ao nível Periférica.

Quadro 22 – Aplicação da tabela de decisão no processo de avaliação

Condições	Regras							
Existe Pertinência?	S	S	S	S	N	N	N	N
Existe Densidade?	S	S	N	N	S	S	N	N
Frequência de uso alta?	S	N	S	N	S	N	S	N
Ações	Decisões							
Conservação permanente	X	X						
Conservação temporária			X		X		X	
Eliminação				X		X		X

Fonte: Elaborado pelo autor.

Para facilitar a análise, as decisões que possuem o mesmo valor foram condensadas em uma coluna (Quadro 23).

As regras com valores coincidentes são mantidas e aquelas que possuem valores variados são representados com um hífen.

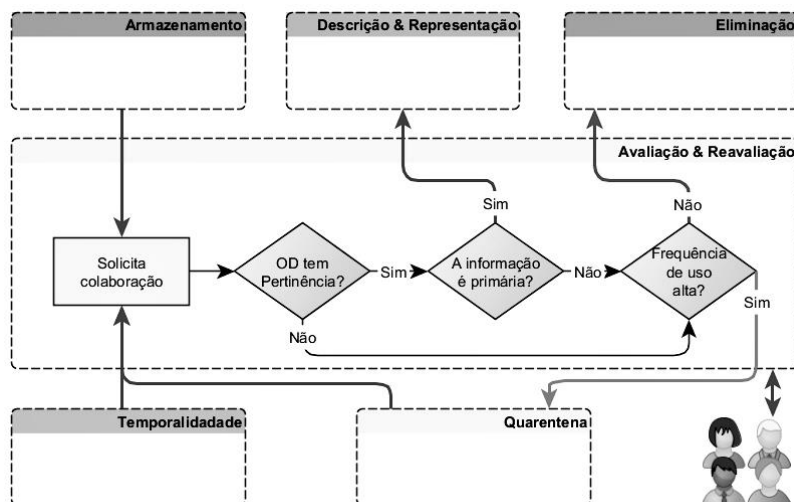
Quadro 23 – Resultado da tabela de decisão do processo de avaliação

Condições	Regras		
Existe Pertinência?	S	-	-
Existe Densidade?	S	-	-
Frequência de uso alta?	-	S	N
Ações	Decisões		
Conservação permanente	X		
Conservação temporária		X	
Eliminação			X

Fonte: Elaborado pelo autor.

A partir do resultado da tabela de decisão, foi elaborado um algoritmo, apresentado no fluxograma da Figura 41.

Figura 41 – Componente Avaliação & Reavaliação



Fonte: Elaborado pelo autor.

A lógica do processo de avaliação conta com três perguntas, que são o núcleo do componente Avaliação & Reavaliação.

Os componentes que operam diretamente com Avaliação e Reavaliação são representados na Figura 41: descrição & representação; quarentena, eliminação; temporalidade; armazenamento e a rede de colaboração.

A denominação original de Silva e Ribeiro (2000, 2004, 2009) dos termos "conservação permanente" e "conservação temporária" foi alterada para "preservação permanente" e "quarentena", respectivamente, por duas razões. Primeiro, para diferenciar o objetivo do *framework*, que é a preservação de objetos digitais visando à preservação do conhecimento e a não a preservação de documentos arquivísticos digitais. Segundo, porque o *framework* proposto prevê a possibilidade de um objeto migrar da quarentena para preservação permanente, diferentemente do documento arquivístico, que se destina à guarda temporária, não é prevista a mudança de destinação para guarda permanente.

4.4 COMPONENTES DAS AÇÕES DE PRESERVAÇÃO

4.4.1 Componente Dependência Tecnológica

A necessidade do uso de dispositivos para consumir a informação armazenada em objetos digitais é a diferença fundamental dos documentos digitais em relação aos documentos em papel.

Diferentemente do documento em papel, o documento digital exige recursos tecnológicos para decodificar os bits em uma representação inteligível a humanos. A falta da tecnologia específica para o documento digital que se deseja visualizar dificulta ou, até mesmo, leva à perda total do conteúdo, como no caso de documento produzido em mídia e *software* descontinuados há muito tempo. É a dependência tecnológica dos objetos digitais.

A maioria dos produtos e serviços do mercado atual foi criada nos últimos cinco anos e há carência de métodos comprovados que garantam: a) que as informações serão preservadas; b) que essas informações serão acessadas usando tecnologias disponíveis; e c) que qualquer informação acessível seja autêntica e confiável (CHEN, 2001).

Ao fazer uso de um ambiente de TIC, uma organização tem dependência de grande quantidade de componentes tecnológicos interligados, fornecidos por inúmeros fornecedores.

No uso cotidiano, essa complexidade passa despercebida, mas tem impacto direto sobre o processo de preservação dos objetos digitais.

O conteúdo digital de uma organização é armazenado em vários computadores, em mídias óticas e magnéticas de variados tipos, e criados em diferentes *softwares* que geram objetos com diferentes formatos. A preservação digital exige a gestão desses componentes a fim de garantir o acesso ao conteúdo.

A gestão do ambiente tecnológico de uma organização é onerada proporcionalmente ao tamanho e à quantidade de componentes tecnológicos utilizados.

A idade do ambiente tem forte impacto, tendendo a um maior legado de formatos de arquivos que leva à necessidade de manter um legado de tecnologia.

O processo de migração do acervo para tecnologias atualizadas também são complexas, considerando que a migração gera perdas e possibilita a adulteração dos objetos, intencionalmente ou não, comprometendo a autenticidade dos objetos. O problema das perdas e alteração fica mais difícil de ser contornado quando percebido, tempos depois da migração, quando a tecnologia originalmente utilizada para a geração do objeto estará ainda mais obsoleta.

Considerando essa complexidade, é possível entender que, no processo decisório de mudança tecnológica numa organização, a decisão de adquirir uma nova tecnologia, geralmente ocorre com entusiasmo, decorrente das novas facilidades e possibilidades que serão proporcionadas ao ambiente. Em oposição, a decisão de desativar uma tecnologia é uma decisão complexa, considerando os riscos e as perdas, frequentemente difíceis de serem inferidos.

O componente Gestão de Dependência Tecnológica visa apoiar o processo de gestão do ambiente em relação ao conteúdo existente na organização.

4.4.1.1 Ontologia DeTec

A dependência tecnológica do objeto digital ficou explicitada em oito classes principais:

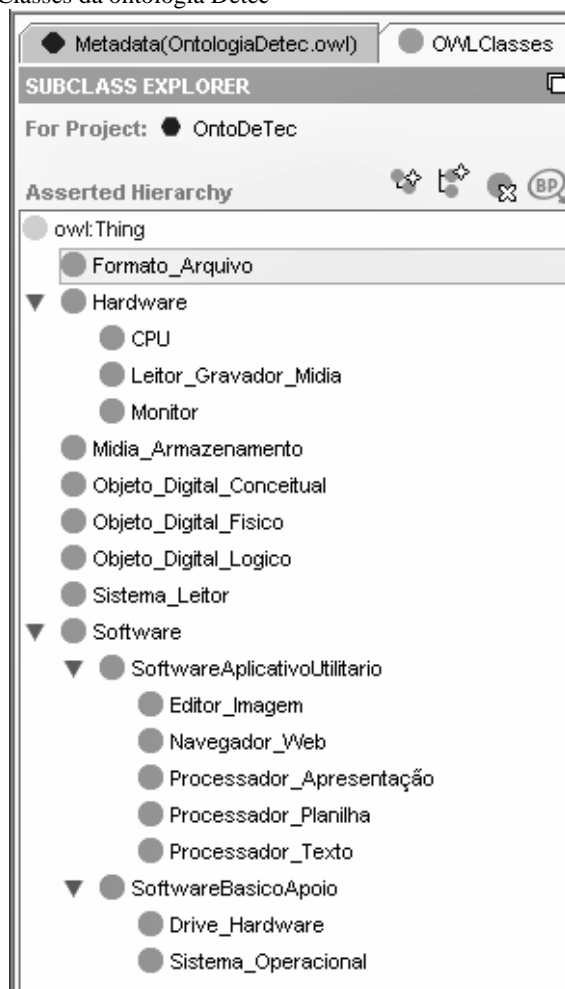
- 1 Formato_Arquivo;
- 2 Hardware;
- 3 Midia_Armazenamneto;
- 4 Objeto_Digital_Conceitual;
- 5 Objeto_Digital_Fisico;
- 6 Objeto_Digital_Logico;
- 7 Sistema_Leitor; e
- 8 Softwaware.

A Figura 42, apresenta as classes na visão da janela de classes do Protégé.

O objeto digital foi representado em três classes, adotando os conceitos propostos por Thibodeau (2005): objeto físico, objeto lógico e objeto conceitual.

A classe Sistema_Leitor é um elemento conceitual, é o agregador de componentes necessários para interpretar o objeto digital lógico para apresentação do objeto digital conceitual. O Sistema_Leitor permite mapear o *hardware*, o *software*, o formato de arquivo e a mídia de armazenamento, logo, um determinado objeto digital deverá, necessariamente, se relacionar com um sistema leitor por meio do formato do arquivo.

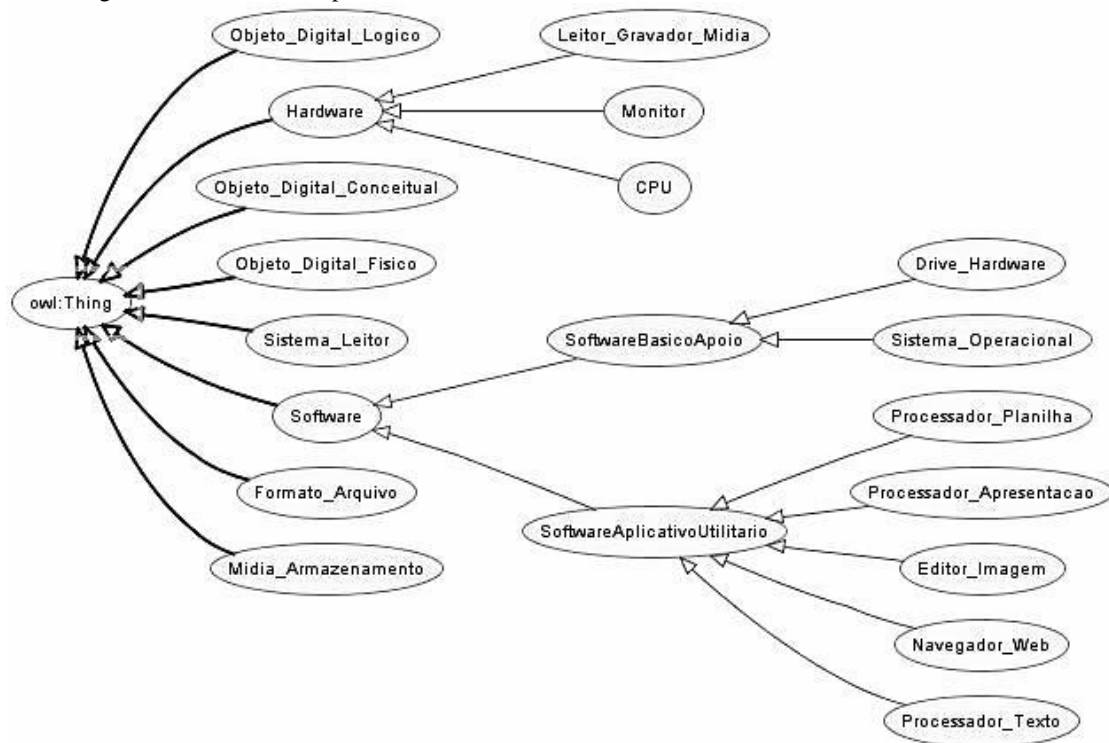
Figura 42- Classes da ontologia Detec



Fonte: Elaborado pelo autor.

A Figura 43 apresenta a visualização da taxonomia da DeTec utilizando o módulo de extensão OWLViz.

Figura 43 – Ontologia DeTec visualizado pelo OWLViz



Fonte: Yamaoka, Gauthier (2012)

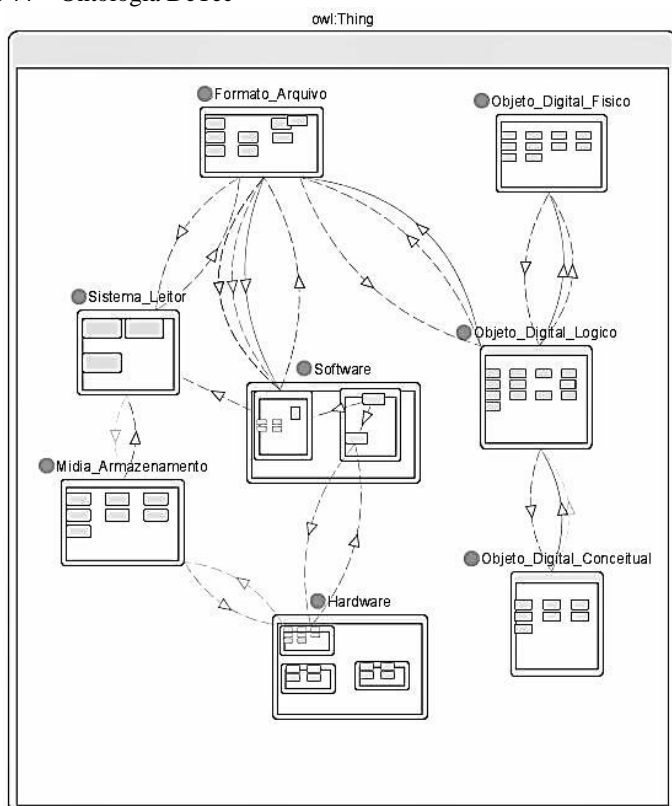
Com a definição das propriedades, isto é, das relações entre as classes, e das restrições, ficou evidente a importância do objeto digital lógico no modelo.

A ontologia assume as seguintes definições:

- um objeto digital conceitual está representado em um ou mais objetos digitais lógicos;
- um objeto digital lógico está armazenado em um (obrigatoriamente) objeto digital físico ou mais;
- um objeto digital lógico tem, necessariamente, um formato de arquivo.

As classes, propriedades e restrições são apresentadas na Figura 44, num diagrama da ontologia DeTec.

Figura 44 – Ontologia DeTec



Fonte: Yamaoka, Gauthier (2012)

O diagrama foi gerado por meio do Jambalaya, que é um módulo de extensão do Protégé, desenvolvida pela universidade de Victoria do Canadá.

É visível na Figura 44 que as maiores quantidades de relacionamentos ocorrem em três classes: formato de arquivo, objeto digital lógico e *software*. A quantidade maior de relacionamentos é um indicativo da importância dessas classes no modelo.

O formato de arquivo especifica a codificação utilizada para o objeto lógico, além de permitir a identificação do *software* a ser utilizado para a apresentação do documento digital (objeto digital conceitual).

O objeto digital lógico detém as informações de codificação e formatação do documento, o que permite a sequência de bits do objeto físico ser recuperado como um objeto conceitual inteligível por humanos.

A ontologia DeTec permite mapear e responder a questões como:

- de qual *hardware* e *software* um determinado documento digital é dependente;
- de qual (is) *software(s)* um determinado *hardware* é dependente;
- de quais *hardwares* e *softwares* uma mídia de armazenamento é dependente;
- numa descontinuidade de tecnologia, qual o impacto e quais mídias de armazenamento e formatos de arquivos requerem migração;
- de qual *software* outro *software* é dependente; etc.

A desativação de um componente tecnológico sem um correto diagnóstico é um grande risco, e a ontologia DeTec possibilita mapear e responder o impacto da desativação de algum componente de tecnologia numa organização, reduzindo o risco de perdas de importantes conhecimentos armazenados em documentos digitais.

Quando um componente tecnológico necessita ser desativado, a organização pode identificar os objetos digitais que serão afetados, podendo planejar a migração desses objetos para formatos atualizados, suportados pelo ambiente tecnológico.

O mapeamento de quais *hardwares* e *softwares* devem ser mantidos e quais podem ser desativados é também um caminho para a racionalização do complexo ambiente de TIC que as organizações possuem, reduzindo custos e riscos.

Com a ontologia DeTec, foi possível perceber que as frequentes afirmações que o documento digital que tem dependência de *hardware* e *software* deve ser reinterpretado, pois:

- é o formato que tem dependência do *software*; e
- é a mídia de armazenamento que tem dependência do *hardware*.

Essa diferenciação conceitual tem impacto na concepção e no desenvolvimento de soluções de preservação digital.

4.4.1.2 Processo de verificação da Dependência Tecnológica

A Figura 45 apresenta um fluxograma do processo de gestão de dependência tecnológica.

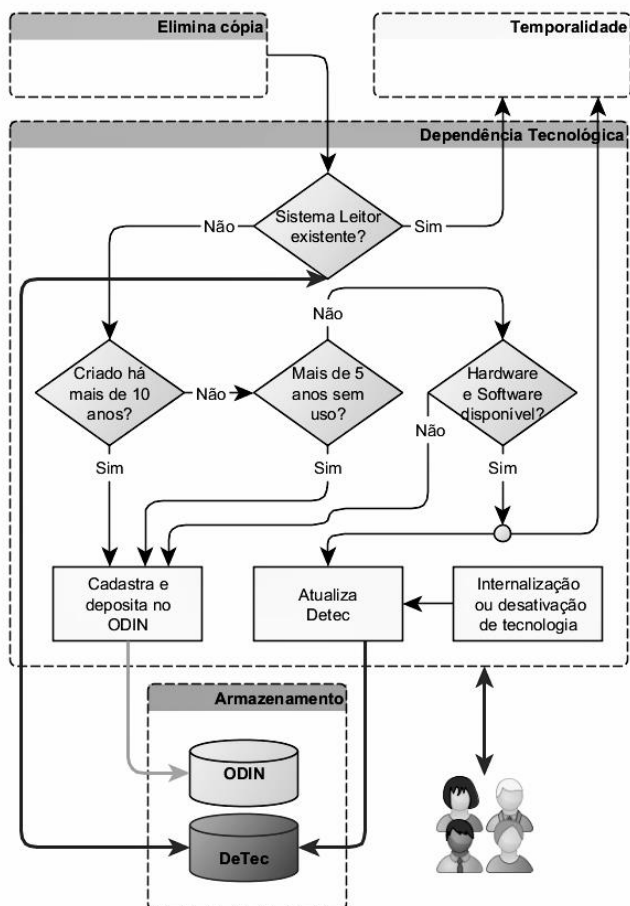
O objeto digital recebido, após a confirmação que é objeto único no *framework*, é verificado se é um objeto que possui Sistema Leitor no DeTec. Se o sistema leitor não existir, significa que o objeto não terá garantia de ser processado e visualizado. Esses objetos sem sistema leitor, caso tenham mais de dez anos de criação¹⁹ ou mais de cinco anos sem uso, são depositados por tempo determinado em um repositório denominado de Objetos Digitais de Identificação Negativa (ODIN). O repositório ODIN foi concebido como segurança adicional a fim de evitar o descarte indevido.

Os objetos criados há menos de dez anos ou que foram utilizados há menos de cinco anos são direcionados para busca de tecnologia que possibilite criação de um sistema leitor. Se o *hardware* e o *software* necessários forem localizados, o sistema leitor é criado e o objeto é encaminhado para etapas subsequentes do *framework*. Se não for possível a criação do sistema leitor, o objeto será depositado no ODIN.

Na prática, um objeto digital depositado no ODIN é um objeto com pouca possibilidade de ser utilizado.

¹⁹ Considerando que, pela Tabela de Temporalidade, 85,56% dos documentos requerem a guarda de, no máximo, 10 anos (seção [Componente Temporalidade](#)).

Figura 45 – Componente Dependência Tecnológica



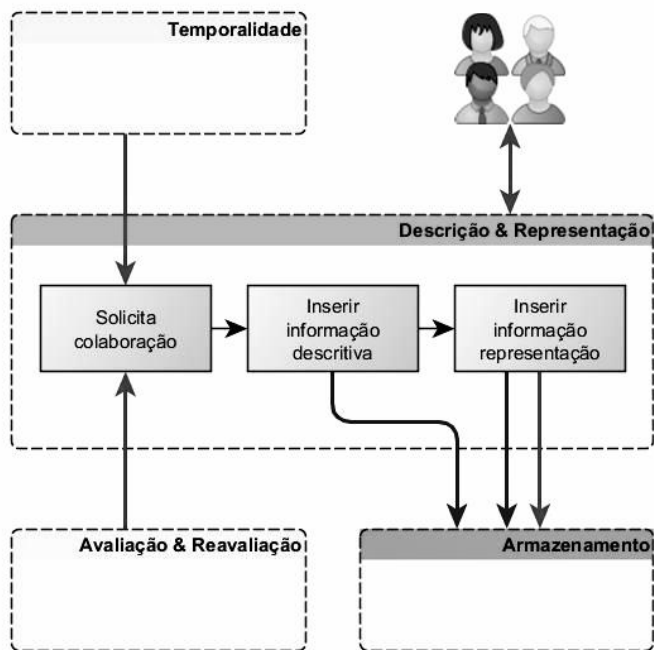
Fonte: Elaborado pelo autor.

A DeTec é, também, acionada para atualização nos eventos de internalização de nova tecnologia ou de desativação de tecnologia pela organização.

4.4.2 Componente Informação Descritiva

Informação de descrição consiste, basicamente, em descrições para apoiar a busca e recuperação dos objetos digitais, ou seja, é a informação que é usada para descobrir qual pacote tem as informações de conteúdo de interesse do usuário (CCSDS, 2012).

Figura 46 – Componentes do grupo Descrição & Representação



Fonte: Elaborado pelo autor.

Metadados são “dados estruturados e codificados que descrevem e permitem acessar, gerenciar, compreender e/ou preservar outros dados do tempo” (ARQUIVO NACIONAL, 2005). Metadados são importantes para criar uma camada que descreve os objetos digitais.

Atualmente, existem vários padrões desenvolvidos para aplicação em algum domínio, a exemplo do metadados para dados geoespaciais, desenvolvido pela *Federal Geographic Data Committee - FGDC*, o *Learning Object Metadata - LOM*, desenvolvido pela *IEEE (Institute of Electrical and Electronics Engineers)* para objetos de aprendizagem, e o *Dublin Core*, um padrão ISO (15836), desenvolvido e mantido pela *Dublin Core Metadata Initiative (DCMI)*.

Existem várias classificações de tipos de metadados. A NISO (2004) identifica os principais tipos de metadados, como:

- Metadados Descritivos, para descrever os recursos com o propósito de identificação e a localização, como, por exemplo, título, autor, palavras-chave;

- Metadados Estruturais, para indicar como os objetos estão compostos, como, por exemplo, como as páginas estão ordenadas em capítulos;
- Metadados Administrativos, para informações, visando apoiar a gestão de recursos, como, por exemplo, a data de criação, tipo de arquivo;
- Metadados para Gerenciamento de Direitos, para questões relacionadas à propriedade intelectual;
- Metadados de Preservação, que apoiam o arquivamento e a preservação de recursos.

O governo brasileiro desenvolveu um padrão específico para o governo eletrônico: o Padrão de Metadados para o Governo Eletrônico (e-PMG), com base no padrão Dublin Core.

O e-PMG conta com 20 elementos com objetivo de facilitar a localização, acesso e gestão dos recursos informacionais (CEGE - COMITÊ EXECUTIVO DE GOVERNO, 2010).

O e-PMG especifica o elemento “Preservação” que conta com os qualificadores, conforme descrito no Quadro 24.

Quadro 24 – Qualificadores do elemento Preservação do e-PMG

Qualificador	Descrição
preservação.formatoOriginal	Nome do formato em que o objeto digital foi inicialmente criado. Ex.: Portable Document Format, Open Document Format, Rich Text Format.
preservação.dependênciaDeHardware	<i>Hardware</i> necessário para acessar o objeto digital.
preservação.dependênciaDeSoftware	<i>Software</i> e versão necessários para acessar o objeto digital.
preservação.DependênciaDeSistemaOperacional	Sistema operacional e versão necessários para acessar o objeto digital.
preservação.resolução	Resolução de imagem em pixels.
preservação.dataDaAção	Data e hora em que uma ação de preservação foi realizada.
preservação.tipoDeAção	A ação de preservação que foi realizada. Valores: conversão, atualização, microfilmagem.
preservação.descriçãoDaAção	Detalhes específicos da ação de preservação, incluindo o estado original, as mudanças realizadas,

	os motivos das mudanças e a autorização para as mudanças.
preservação.ferramentasDaAção	Ferramentas utilizadas na ação de preservação.
preservação.proximaAção	Próxima ação ou avaliação de preservação pela qual o objeto digital deve passar.
preservação.dataDaProximaAção	A data programada para a execução da próxima ação.

Fonte: Baseado em Comitê Executivo de Governo Eletrônico (2010)

Atualmente, é consenso que os conteúdos digitais devem estar acompanhados de metadados para possibilitar que, no futuro, sejam acessados e compreendidos (SAYÃO, 2010), para isso, são necessários a adoção e o uso de vários padrões.

Permitir a codificação dos diversos tipos de metadados e vinculá-los aos objetos correspondentes é o objetivo do Metadata Encoding & Transmission Standard (METS) (METS EDITORIAL BOARD, 2013).

Para o componente descrição do framework, o padrão METS apresenta-se como uma solução, considerando que a linguagem utilizada é o XML, que também é um padrão aberto e internacional, além de ser interpretável por humanos.

O METS permite o uso de esquemas externos aprovados (endossados), que são aqueles que representam a necessidade de uma comunidade e que passa por uma revisão visando ao funcionamento efetivo no contexto do METS. Entre os esquemas externos endossados, estão o Dublin Core, LOM, FGDC, MARC, e o PREMIS.

4.4.3 Componente Informação de Representação

Informação de representação é qualquer informação necessária para compreender e reproduzir o objeto digital e os metadados associados. São as informações sobre a estrutura e a semântica que permitem converter as cadeias de bits em informação significativa.

A informação estrutural interpreta os bits, especifica o formato e a descrição do ambiente de *hardware/software* necessário para acessar os dados e a informação semântica que fornece significados adicionais para os dados (OCLC/RLG WORKING GROUP ON PRESERVATION METADATA, 2002) . Por exemplo, a informação estrutural identifica que a cadeia de bits é um texto em ASCII e a informação semântica indica que o texto está em um determinado idioma.

O modelo de referência OAIS estabelece que a informação de representação pode ser substituída pelo *software* de acesso ao objeto digital.

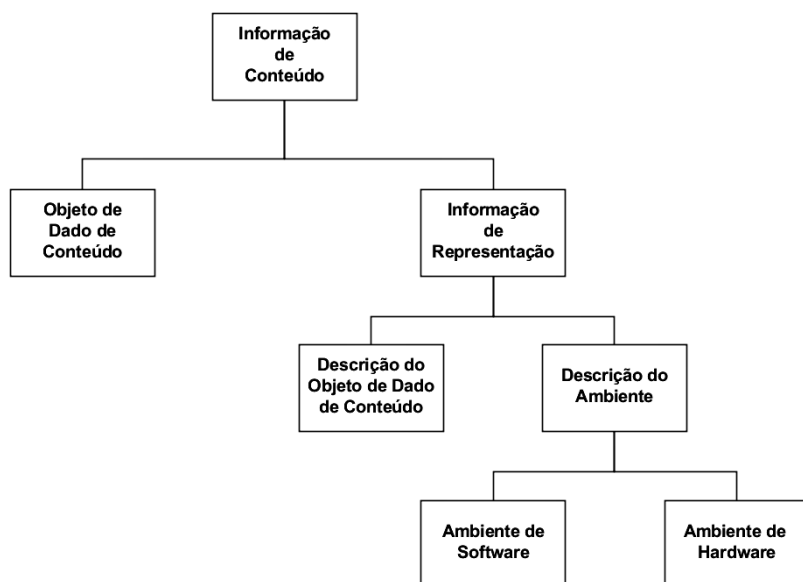
Para proporcionar maior possibilidade de preservação de longo prazo, o *framework* propõe manter ambos os recursos: a informação de representação e o *software* de acesso.

A Figura 47 ilustra o pacote de informação de conteúdo, que é o Objeto de Dado de Conteúdo e sua informação de representação.

A descrição do objeto de dado do conteúdo deve detalhar as características do objeto de dado de conteúdo, necessárias para processar e entender o seu conteúdo.

O componente Descrição do ambiente deve descrever o ambiente de *hardware* e *software* capaz de processar ou exibir o conteúdo do objeto de dado (OCLC/RLG WORKING GROUP ON PRESERVATION METADATA, 2002).

Figura 47 – Pacote de Informação de Conteúdo



Fonte: Adaptado de OCLC/RLG Working Group on Preservation Metadata (2002).

No *framework*, a existência da ontologia DeTec resolve o problema da descrição de ambiente. A classe "Sistema Leitor" da DeTec reúne as informações referentes.

No tocante à descrição do objeto de dado de conteúdo, a escolha de formato de arquivo com determinadas características facilita o registro da descrição, como é o caso do formato ODF apresentado na seção 2.11, página 113.

4.4.4 Atualização

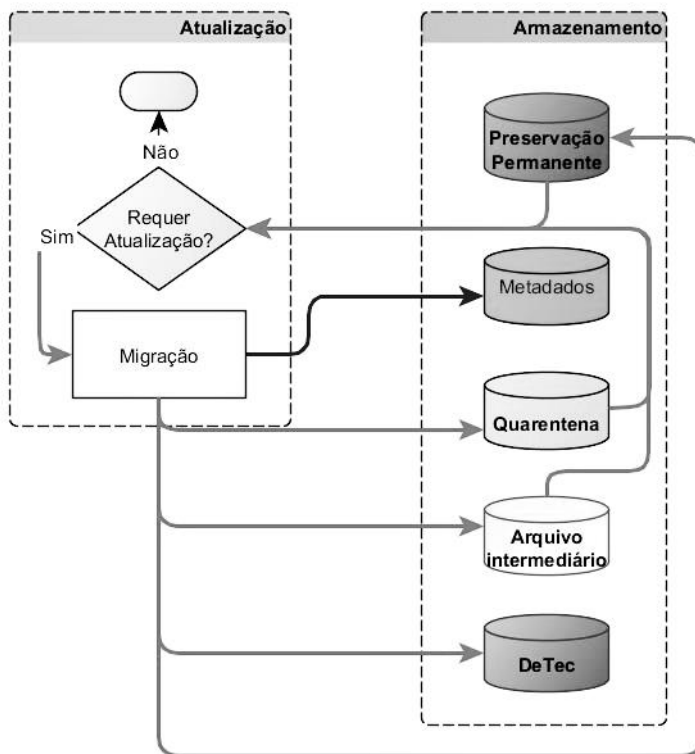
Assim como previsto no ciclo de vida de curadoria digital do DCC (seção 2.6 - Curadoria Digital, página 96), o *framework* faz uso da estratégia de migração dos objetos digitais.

A migração é uma forma de transpor a obsolescência tecnológica, transferindo ou convertendo o objeto digital para tecnologia atualizada.

O objetivo da migração é a preservação do objeto digital conceitual, ou seja, a manutenção da capacidade de processar e visualizar o conteúdo do objeto digital com a tecnologia disponível.

A Figura 48 apresenta o processo geral do componente Atualização.

Figura 48 – Componente Atualização



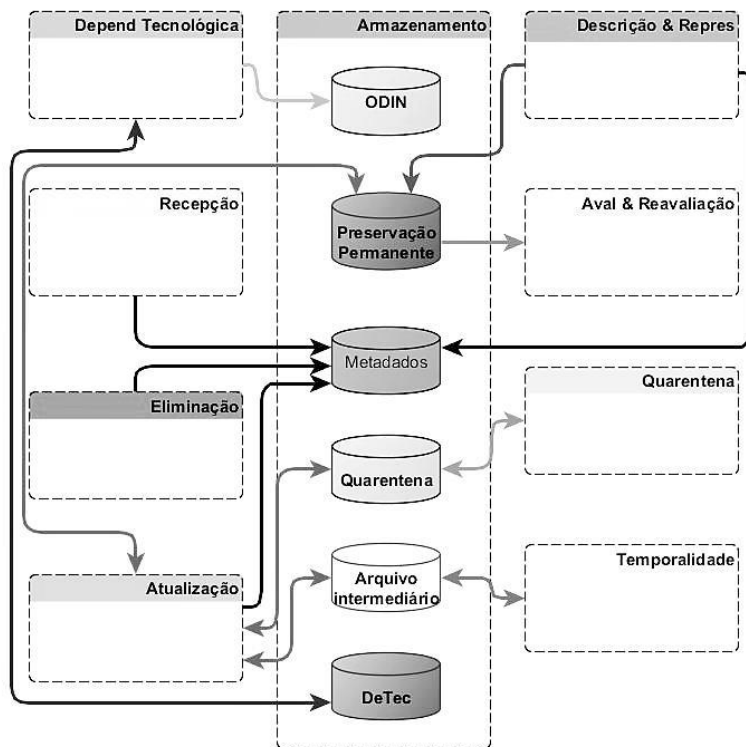
Fonte: Elaborado pelo autor.

Os três repositórios de objetos digitais (Preservação Permanente, Quarentena e Arquivo Intermediário) são monitorados quanto à necessidade de atualização. Todas as migrações realizadas nos objetos digitais devem ser registradas na DeTec e nos Metadados.

4.5 ARMAZENAMENTO

O componente Armazenamento tem a função de garantir a persistência dos objetos digitais pelo tempo que for necessário. Para tanto, faz interação com praticamente todos os componentes do *framework*, como pode ser observado na Figura 49.

Figura 49 – Componente Armazenamento



Fonte: Elaborado pelo autor.

Conceitualmente, são quatro repositórios com o objetivo de armazenar objetos digitais (ODIN, Preservação Permanente, Quarentena e Arquivo Intermediário), um repositório de Informação de Descrição (Metadados) e um repositório para gestão da Dependência Tecnológica (DeTec).

Como componente crítico para a preservação digital, torna-se fundamental que o componente armazenamento tenha total conformidade com o conceito de repositório digital confiável desenvolvido pelo grupo de trabalho do RLG com a OCLC, denominado de *RLG/OCLC Working Group on Digital Archive Attributes*.

Um Repositório Digital Confiável é aquele cuja missão é proporcionar o acesso confiável e de longo prazo dos recursos digitais, para isso, deve (RLG/OCLC, 2002):

- aceitar a responsabilidade pela manutenção de longo prazo dos recursos digitais;
- ter uma estrutura organizacional que suporte a viabilidade de longo prazo do repositório e também a informação digital sobre a qual tem responsabilidade;
- demonstrar responsabilidade fiscal e sustentabilidade;
- projetar os sistemas de acordo com convenções e padrões para assegurar a gestão, o acesso e a segurança dos materiais depositados;
- estabelecer metodologias para avaliação dos sistemas que considerem as expectativas de confiabilidade da comunidade;
- considerar, no desempenho de suas responsabilidades de longo prazo, os depositários e os usuários; e
- ter políticas, práticas e desempenho que sejam auditáveis e mensuráveis.

Portanto, para que um repositório digital seja confiável, é necessário muito mais do que uma infraestrutura tecnológica moderna e robusta. Exige o estabelecimento de política, processos e uma equipe que dê sustentação.

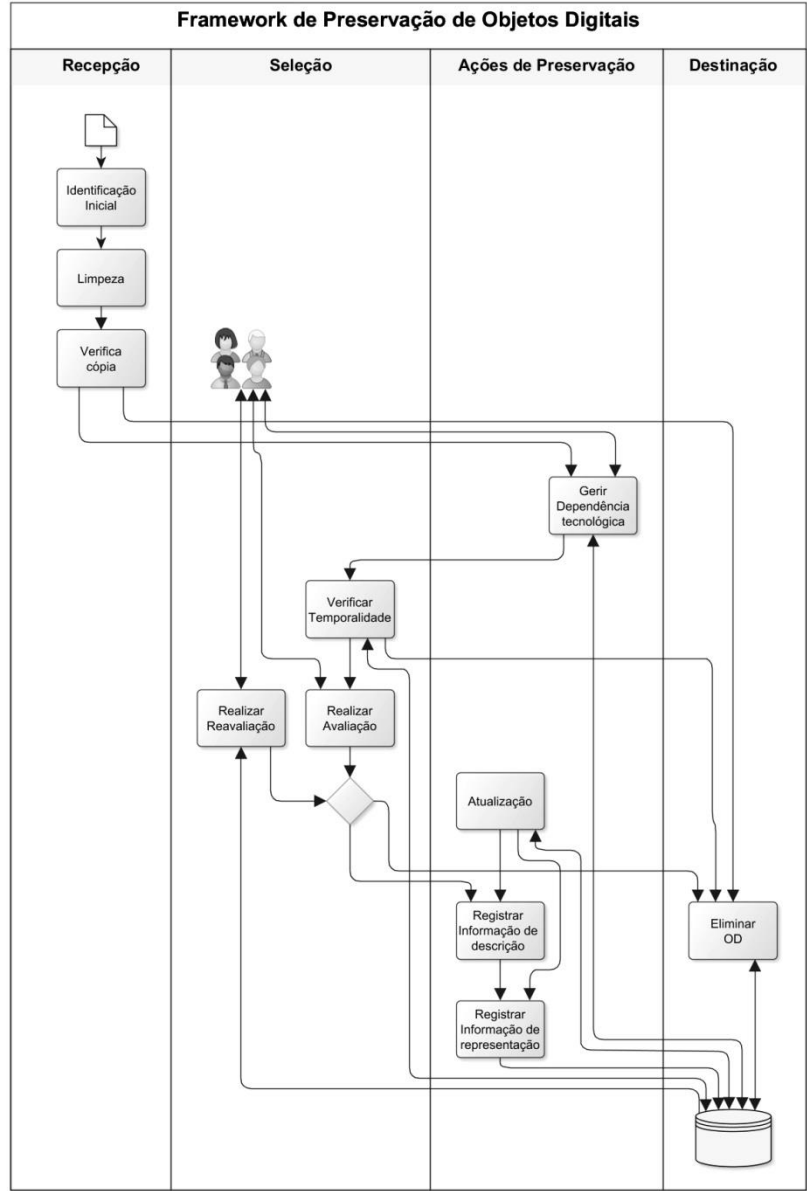
4.6 FRAMEWORK – VISÃO GERAL

Para permitir a leitura completa do *framework*, a Figura 50 apresenta os principais componentes e seus relacionamentos dispostos nos quatro principais grupos: Recepção, Seleção, Ações de Preservação e Armazenamento.

No apêndice D, a

Figura 57 apresenta a visão geral detalhada.

Figura 50 – Framework Visão Geral Resumida



Fonte: Elaborado pelo autor.

5 VERIFICAÇÃO E AJUSTES DO FRAMEWORK

5.1 ANÁLISE DAS RESPOSTAS

As respostas das entrevistas estão agrupadas em seis tópicos, de acordo com o roteiro da entrevista.

5.1.1 Aplicabilidade do framework nas organizações

Todos os entrevistados consideraram o *framework* uma proposta aplicável.

Dois entrevistados manifestaram a preocupação que, devido à complexidade e ao possível custo para desenvolvimento e operacionalização do *framework*, seria necessário ser implementado em organizações com porte, condições financeiras e estruturais.

O custo para a operacionalização é uma barreira, mas, de acordo com um dos entrevistados, no setor governamental, por exemplo, é uma necessidade inquestionável, pois a preservação dos documentos pelo tempo necessário é uma forma de garantir os direitos do cidadão. A necessidade também é decorrente de imposições legais.

Na percepção dos entrevistados, atualmente, as organizações não contam com um processo de preservação de objetos digitais.

5.1.2 O processo de seleção dos objetos digitais

Cinco dos seis entrevistados se manifestaram positivamente ao processo de seleção. Todos reconhecem que é inviável guardar tudo.

Dois entrevistados manifestaram preocupação quanto a um objeto considerado sem importância no momento da avaliação vir a ser considerado importante no futuro, o que foi denominado de falso negativo por um dos entrevistados.

A preocupação da possibilidade de avaliação sem a imparcialidade necessária também foi manifestada por um entrevistado, principalmente em documentos que sejam politicamente sensíveis. Porém, reconheceu que é um problema humano, que transcende ao *framework*.

Um entrevistado manifestou que:

- a) a introdução da avaliação dos documentos não arquivísticos é um ganho maravilhoso;

- b) a reavaliação dos documentos arquivísticos proporciona uma segurança adicional, pois há sempre a insegurança dos avaliadores quanto à eliminação de documentos;
- c) que o processo permite a seleção mais controlada, mais consequente e possibilita a eliminação necessária, com segurança.

5.1.3 Grupo ações de preservação

Todos os entrevistados concordaram que as ações de preservação são importantes. São ações que estão ausentes nos sistemas atuais. Sem as ações de preservação o que são preservados são somente os bits. Com o tempo, perde-se o acesso ao objeto e um objeto inacessível é o mesmo que não o ter.

5.1.4 Identificação dos pontos fortes do framework

Os seguintes pontos fortes do *framework* foram identificados pelos entrevistados:

- a) a abrangência, pois supre todo o ciclo do processo de preservação de informação e conhecimento;
- b) os processos que estão bem descritos;
- c) o fato dele ser um *framework* voltado para documento digital – o que não existe hoje nas organizações;
- d) o processo de monitorar e superar a obsolescência tecnológica;
- e) o processos de avaliação e seleção;
- f) as ações de preservação;
- g) a ontologia de dependência tecnológica;
- h) o processo de colaboração;
- i) a estratégia da quarentena, como salvaguarda para o falso negativo;
- j) o processo de reavaliação que permitirá conhecer cada vez melhor o conteúdo da organização.

5.1.5 Identificação dos pontos para Melhorias

Pontos que podem ser melhorados na visão dos entrevistados:

- a) Como o processo de migração não é infalível, como garantir que um objeto foi migrado corretamente? Como testar e garantir transformações confiáveis?

- b) inclusão do processo do gestor do *framework*;
- c) detalhar mais alguns processos, como, por exemplo, a quarentena;
- d) se possível, tornar o *framework* imune às questões de parcialidade no processo de seleção;
- e) o *framework* poderia fazer referência à questão do certificado digital;
- f) a questão da identificação da cópia. Dois documentos iguais, se pertencerem a duas instituições (dois fundos), são documentos diferentes, a exemplo de um ofício que é enviado de uma instituição para outra. O documento enviado e o documento recebido, embora sejam idênticos, são documentos distintos. O documento é único para cada atividade que participa.

5.1.6 Grau de satisfação

O grau de satisfação dos entrevistados foi elevado. Na escala de zero a 10, um entrevistado respondeu como "certamente acima de 8", dois entrevistados atribuíram grau 9 e três entrevistados atribuíram grau de satisfação 10.

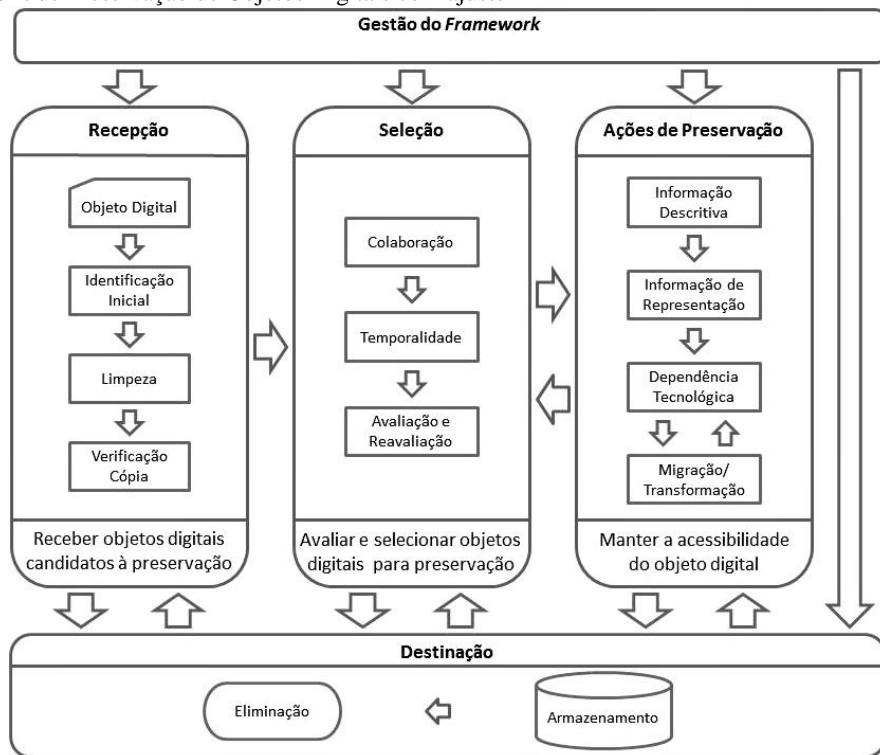
5.2 AJUSTES REALIZADOS NO FRAMEWOK

Com base nas sugestões de melhorias e aprimoramento apontados pelos entrevistados, foram realizados ajustes no *framework* visando corrigir ou minorar as carências e introduzir os aprimoramentos possíveis, limitado pelo escopo deste trabalho.

5.2.1 Gestor do framework

A Gestão do Conhecimento requer a participação ativa de todas as áreas da organização, porém, indiscutivelmente, um processo complexo como a preservação digital iria requerer a presença do gestor. A Figura 51 apresenta o *framework* introduzindo o elemento Gestão.

Figura 51 – *Framework* de Preservação de Objetos Digitais com ajuste



Fonte: Elaborado pelo autor.

Considerando o objetivo de preservar o conhecimento armazenado nos objetos digitais, um candidato natural para coordenar o processo de gestão seria a área responsável pela gestão de conhecimento na organização, porém, a operacionalização exige a formação de uma equipe interdisciplinar com a participação de especialistas em tecnologia da informação, gestão documental, bibliotecas e especialistas da área-fim da organização.

5.2.2 Parcialidade no processo de seleção

O risco de ocorrer avaliações sem a isenção necessária é uma possibilidade. No entanto, é possível prevenir e minorar essas ocorrências com algumas medidas:

- processo colaborativo aberto, com o automonitoramento realizado pela própria rede de colaboração e a comunidade usuária;
- instituição da figura de moderadores para traçar ações de correção caso alguma avaliação seja díspar em relação às outras, com o apoio de *software* que ajude a identificação dessas ocorrências.

5.2.3 Objetos assinados com certificado digital

A garantia da autenticidade do conteúdo é um dos pontos críticos na preservação digital. O uso dos certificados digitais tem crescido visando a essa garantia. Porém, a tecnologia que permite a inserção das assinaturas digitais nos documentos também são vítimas da obsolescência tecnológica e dificultam ou impossibilitam a migração do objeto digital para novas versões e formatos, em virtude do seu objetivo e funcionalidade inerente (DURANTI; PRESTON, 2008; INTERPARES PROJECT, 2007). Por este motivo, o Projeto InterPARES recomenda desanexar a assinatura digital do documento e registrar essa informação nos metadados, indicando que o documento originalmente estava assinado com o uso de certificado digital (INTERPARES PROJECT, 2007).

No *framework*, essa operação poderá ser realizada na recepção do objeto digital, especificamente, no processo de limpeza do objeto.

5.2.4 Avaliação com resultado falso negativo

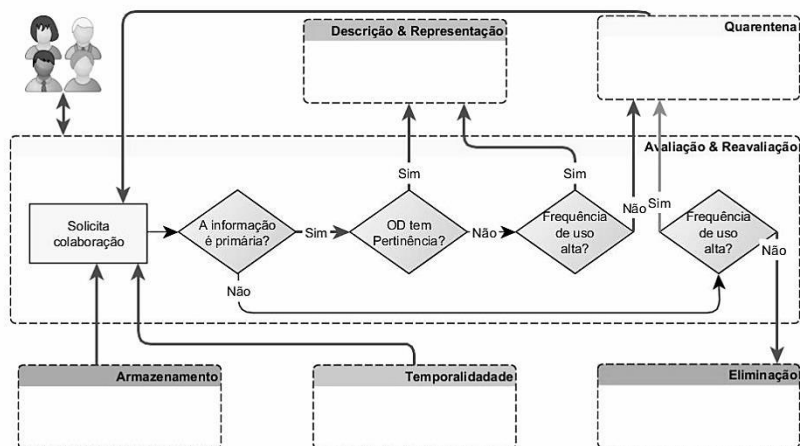
Os especialistas em Gestão do Conhecimento entrevistados manifestaram que um objeto com conhecimento poderia ser considerado descartável no momento da avaliação e, no futuro, se tornar um conhecimento importante.

O framework já prevê o processo de quarentena visando proporcionar maior segurança no processo de descarte. Após novo estudo do processo de avaliação e reavaliação, com o objetivo de proporcionar maior segurança no processo de avaliação, foi introduzida mais uma alteração na metodologia proposta originalmente por Silva e Ribeiro (2000, 2004, 2009).

Como o objetivo original da metodologia é a avaliação da informação arquivística, o critério "pertinência" determina se o objeto tem a guarda definitiva ou não. A alteração introduzida, visando à preservação do conhecimento, é a atribuição de um peso maior para o critério "densidade", ou seja, a informação primária passa a ter maior relevância no processo de avaliação.

Com essa alteração, o processo de avaliação muda consideravelmente, como pode ser observado na Figura 52.

Figura 52 – Processo do Componente Avaliação & Reavaliação Ajustado



Fonte: Elaborado pelo autor.

Com o ajuste no processo, se uma informação é primária e tem ou teve frequência de uso, o objeto é destinado à preservação permanente.

Na metodologia original de Silva e Ribeiro (2000, 2004, 2009), nesses casos, o objeto iria para guarda temporária (ou para a quarentena, conforme a proposta inicial do *framework*).

Com este ajuste, o *framework* passa a dar relevância para um objeto que contém um conteúdo primário (não derivado).

Essa alteração, aparentemente simples, possibilita introduzir mudanças importantes no processo de Gestão do Conhecimento e inovação da organização.

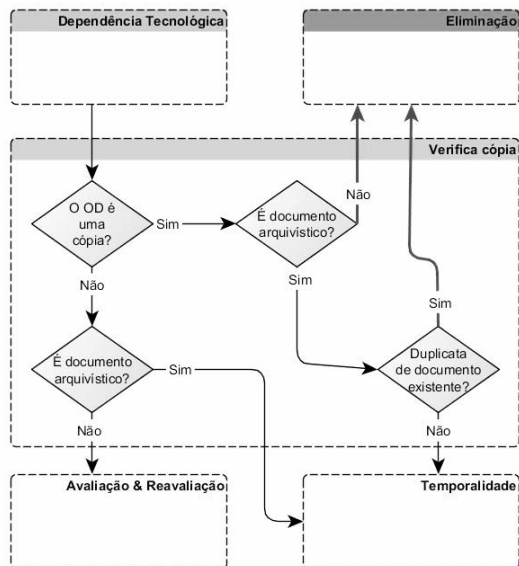
Um objeto com conteúdo primário, que tem uso frequente e não possui pertinência em relação à missão e objetivo da organização, pode ser um indicador da necessidade de revisão dos seus objetivos ou oportunidade de inovação na organização.

5.2.5 Documentos idênticos como documentos distintos.

A possibilidade de dois objetos idênticos, se forem documentos arquivísticos, serem documentos distintos, exigiu também uma revisão no componente "verifica cópia".

O componente foi ajustado para explicitar a identificação se dois objetos idênticos podem ser documentos arquivísticos distintos ou não. A Figura 53 apresenta o processo ajustado.

Figura 53 – Componente Verificação de Cópia Ajustado



Fonte: Elaborado pelo autor.

A explicitação é justificada considerando que os documentos arquivísticos constituem um subconjunto da totalidade do conteúdo, portador de conhecimento codificado, de uma organização.

O *framework* tem como premissa não interferir nas operações do Arquivo.

5.2.6 Ajuste do diagrama geral do framework

O diagrama com a visão geral do *framework* com os ajustes realizados após as entrevistas é apresentado na Figura 58, no Apêndice E.

6 CONCLUSÕES

Um problema real, emergente e crítico para as pessoas, para as organizações e, em última análise, para a humanidade foi o motivador deste trabalho.

A TIC possibilitou crescimento exponencial da produção e disseminação da informação, porém, num processo autofágico, tornou-se vítima da sua própria evolução, armazenando informações e conhecimentos em objetos digitais que, em pouco tempo, ficam degradados pela obsolescência tecnológica.

Considerando que praticamente a totalidade de novas informações atualmente é criada e armazenada em objeto digital, medidas urgentes são necessárias para possibilitar o acesso futuro a eles, caso contrário, parte da memória será perdida para sempre. Essa preocupação ficou presente nas respostas dos entrevistados no processo de validação do *framework* proposto.

Durante o desenvolvimento deste trabalho, observou-se que alguns fundamentos, princípios e processos em uso foram desenvolvidos no contexto em que o papel era o suporte para registro da informação e do conhecimento. É o de alguns princípios da Arquivologia.

As diferenças na forma utilizar, dar ou emprestar um arquivo de computador é muito diferente de um documento ou livro em papel. Se não for protegido, o cedente pode fornecer cópias idênticas da obra que tem posse, sem dificuldades, porém, se for um arquivo protegido até o empréstimo será difícil. Portanto, processos, regras de negócio e legislação ainda estão em processo de adaptação.

O princípio da proveniência, também conhecido como o princípio de respeito aos fundos, que é um princípio basilar da Arquivologia, estabelece que dois documentos idênticos podem ser documentos distintos se pertencentes a fundos diferentes. Na era digital, esse princípio tem sido colocado em xeque por estudiosos como Rousseau e Couture (1998).

No caso do documento arquivístico digital, uma adequada arquitetura do sistema poderá, se for o caso, manter somente uma cópia física do documento digital e recuperá-lo, logicamente, em diferentes fundos. Em suma, não haveria a obrigatoriedade de manter várias cópias do mesmo objeto digital.

Exemplos como esses ilustram a necessária revisão nos conceitos e modelos mentais estabelecidos pelo paradigma do objeto físico.

Este trabalho teve como preocupação a preservação do conhecimento armazenado em objetos digitais, ou seja, a preservação

dos elementos presentes ou manifestados no mundo 1 de Popper, em que fica registrado parte do conhecimento explícito das organizações. A Figura 54 situa e ilustra a visão de mundo adotada neste trabalho. Os processos e as tecnologias visam manter o acesso permanente aos elementos do mundo 1 para preservar o conhecimento do mundo 3, o conhecimento sem o conhecedor.

Figura 54 – Conhecimento e Objetos Digitais e Três Mundos de Popper

	Conhecimento	FRBR	Classes Objetos Digitais
Mundo 1 - Material		Item	Objeto Físico
		Manifestação	Objeto Lógico
			Objeto Conceitual
Mundo 3 Ideias	Explícito	Expressão	
		Obra	
Mundo 2 Mental	Tácito		

Fonte: Elaborado pelo autor.

A preservação do conhecimento codificado é um desafio pela própria natureza do objeto a ser preservado.

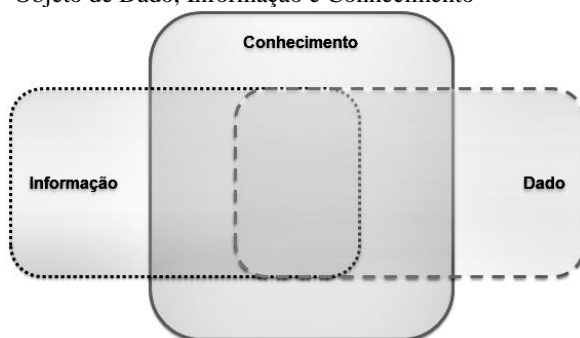
O objeto digital pode ser de um simples arquivo de texto sem formatação, do tipo txt, a objetos mais complexos, como uma planilha eletrônica repleta de fórmulas e imagens ou, até mesmo, um banco de dados inteiro. Como os objetos digitais possuem propriedades significativas, ou seja, conteúdo, contexto, aparência, estrutura e comportamento, presume-se que pode ter conhecimento incorporado em qualquer objeto digital.

Um banco de "dados", por exemplo, armazena dados, informações e conhecimento. É um repositório de dados, possui informações de descrição, como tipo de dado e formato lógico do dado, e possui conhecimento incorporado no próprio projeto do banco de

dados, como a definição dos relacionamentos entre tabelas, os procedimentos e gatilhos²⁰. Portanto, um banco de dados, além dos dados, é um repositório do conhecimento, de como um determinado banco foi projetado e utilizado. É potencialmente um repositório de *formal knowledge*, *instrumentalities* e *contingent knowledge*, da categorização de Fleck (1997).

A Figura 55 que representa a relação entre os objetos de dado, informação e conhecimento, com esse entendimento, ou seja, com a intenção de evidenciar a possibilidade de um único objeto digital ter incorporado o dado, a informação e o conhecimento simultaneamente. A representação na forma de pirâmide DIK não manifesta essa condição e pode, eventualmente, induzir ao pensamento de que um objeto pode ser pertencer somente a uma classe: ou a dado, ou a informação ou a conhecimento.

Figura 55 – Objeto de Dado, Informação e Conhecimento



Fonte: Elaborado pelo autor

O V epistemológico de Gowin, apresentado de forma parcial no início do desenvolvimento deste trabalho na Figura 1 com o domínio conceitual, as questões básicas e o evento planejado, é representado de forma completa na Figura 56, com a síntese do resultado deste estudo.

²⁰ Conhecidos como *triggers* de banco de dados.

Figura 56 – V epistemológico de Gowin completo



Fonte: Elaborado pelo autor.

A pergunta de pesquisa que caracteriza o problema deste trabalho é: "Quais componentes são basilares em um *framework* de preservação de conhecimento codificado armazenados em objetos digitais, considerando o volume e a mutabilidade constante da tecnologia?". No V epistemológico de Gowin, a pergunta foi subdividida em duas questões básicas:

- 1ª) O que é necessário para viabilizar a preservação do conhecimento armazenado em objetos digitais?
- 2ª) Como selecionar objetos digitais que devem receber prioridade na preservação do conhecimento?

A diferença fundamental entre o papel e o digital para acesso ao seu conteúdo é que, no segundo, é requerido o intermédio de dispositivos tecnológicos. Consequentemente, a preservação do conhecimento codificado incorporado nos objetos digitais depende da manutenção desses objetos para mantê-los compatíveis com as tecnologias disponíveis; logo, as ações de preservação composta pelos componentes gestão da dependência tecnológica dos objetos digitais, atualização tecnológica, e a descrição e representação dos objetos digitais, componentes estruturantes do *framework* proposto. No V de Gowin, duas asserções de valor respondem a primeira questão básica:

- A preservação do conhecimento armazenado em objetos digitais é possível, desde que o objeto e a tecnologia que possibilita o seu acesso estejam alinhados;
- A gestão da dependência tecnológica é parte integrante do processo de preservação digital.

O grande volume torna o processo de gestão e atualização dos objetos digitais complexo e oneroso. Portanto, reduzir esse volume, selecionando o mais importante, torna-se uma importante ação a ser considerada.

Todo processo que visa à seleção é frequentemente cercado de polêmicas e dúvidas, no entanto, em várias situações, a impossibilidade de manter ou contemplar a totalidade leva à necessidade de selecionar, com o uso de alguma forma de avaliação.

No caso dos objetos digitais, a seleção dos objetos relevantes para o futuro pode ser o fator que viabiliza ou inviabiliza a preservação digital de longo prazo. A alternativa que é a preservação da totalidade, quicá possível, implica num custo muito superior, e também na dificuldade em gerir e localizar o recurso desejado.

Em razão do exposto, o *framework* propõe o processo de avaliação e seleção de objetos digitais como outro pilar da preservação

digital de longo prazo. No V de Gowin, as seguintes asserções respondem a segunda questão básica:

- Asserção de conhecimento: A seleção criteriosa dos objetos digitais poderá ser realizada com base na avaliação de três critérios: verificar se o conteúdo é primário; se é conteúdo pertinente ao objetivo da organização; e se a frequência de uso foi alta;
- Asserção de valor: A seleção de objetos digitais visando à preservação do conhecimento poderá ser realizada com segurança se os critérios de avaliação forem claramente estabelecidos.

6.1 TRABALHOS FUTUROS

A curadoria e preservação digital visando à preservação do conhecimento e da memória é um campo de estudo recente, que ainda requer muito empenho e investimentos.

No desenvolvimento deste estudo, ficou patente essa necessidade, desde a revisão de literatura, no acompanhamento das notícias, em conversas informais e na entrevista para verificação do *framework* proposto.

A evolução desta proposta para um sistema informação ainda irá requerer estudos aprofundados, em vários aspectos, pois a efetivação da preservação digital exige o repensar nos modelos e processos em uso.

Como recomendação de trabalho futuro, destacam-se alguns, os quais são decorrentes de reflexões e conjecturas realizadas no decorrer do desenvolvimento deste trabalho.

Observou-se o potencial que o *framework* possui, principalmente no tocante ao processo de seleção de objetos relevantes para a preservação de longo prazo. Estudo visando à aplicação do processo de seleção proposto poderá permitir o aprimoramento do processo e experimentos que possibilitarão apurar números efetivos da proporção preservação *versus* eliminação de objetos nas organizações.

O resultado do processo de seleção baseado em trabalho colaborativo irá também permitir o estudo de diferentes abordagens e enforques visando aprimorar e acompanhar a geração de novos conhecimentos baseados no processo decisório dos colaboradores.

Recomenda-se, também, o estudo para verificar se a implementação do *framework* permitirá a gradativa redução do esforço humano, com base de conhecimento que se formará e também no

gradativo enriquecimento da informação de descrição dos objetos digitais.

Finalmente, considerando que os atuais sistemas de informação não contemplam processos e funcionalidades para a preservação dos objetos digitais e a persistência da visão de documentos, recomendam-se estudos visando à revisão e ao aprimoramento do processo de concepção e implementação dos sistemas de informação.

Em qualquer estudo futuro, é fundamental a adoção da premissa de que o processo de preservação digital tem início na criação do objeto digital, portanto, o criador do objeto é também responsável pela preservação do conhecimento codificado e incorporado nos objetos digitais.

REFERÊNCIAS

- ABBOTT, D. **What is digital curation?** Edinburgh: DCC, 2008.
Disponível em: <<http://hdl.handle.net/1842/3362>>. Acesso em: 30 nov. 2013.
- ALAVI, M.; LEIDNER, D. E. Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. **MIS Quarterly**, v. 25, n. 1, p. 107–136, 2001.
- AMARDEILH, F. Semantic Annotation and Ontology Population. In: CARDOSO, J.; LYTRAS, M. D. (Eds.). **Semantic Web Engineering in the Knowledge Society**. Hershey/New York: Information Science Reference, 2009. 135–160 p. .
- AMORETTI, M. S. M.; TAROUÇO, L. M. R. Mapas Conceituais: modelagem colaborativa do conhecimento. **Informática na Educação: Teoria e Prática**, v. 3, n. 1, 2000. Disponível em: <<http://seer.ufrgs.br/InfEducTeoriaPratica/article/view/6412/3854>>. Acesso em: 5 out. 2011
- ARQUIVO NACIONAL. **Classificação, temporalidade e destinação de documentos de arquivo**; relativos às atividades-meio da administração pública. Rio de Janeiro: Arquivo Nacional, 2001. 156 p. .
- ARQUIVO NACIONAL. **DICIONÁRIO BRASILEIRO de Terminologia Arquivística**. Rio de Janeiro: Arquivo Nacional, 2005. 232 p. .
- ARQUIVO NACIONAL. **Tabela de temporalidade e destinação de documentos de arquivo relativos às atividades-fim das instituições federais de ensino superior - IFES**. Rio de Janeiro: Arquivo Nacional, 2011. 49 p. Disponível em: <<http://www.siga.arquivonacional.gov.br/cgi/cgilua.exe/sys/start.htm?sid=62>>. Acesso em: 18 abr. 2013
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 15472: Sistemas espaciais de dados e informações - Modelo de referência para um sistema aberto de arquivamento de informações**. Rio de Janeiro: [s.n.]. 95 p. .

ATHERTON, J. From Life Cycle to Continuum: Some Thoughts on the Records Management - Archives Relationship. **Archivaria**, n. 21, p. 43–51, 1985.

AUER, S. RapidOWL: A Methodology for Enabling Social Semantic Collaboration. In: CARDOSO, J.; LYTRAS, M. (Eds.). **Semantic Web Engineering in the Knowledge Society**. Hershey/New York: Information Science Reference, 2009. 267–289 p. .

BALL, A. **Briefing Paper : the OAIS Reference Model**. Bath, UK : UKOLN - University of Bath, 2006. Disponível em: <<http://www.ukoln.ac.uk/projects/grand-challenge/papers/oaisBriefing.pdf>>. Acesso em: 25 fev. 2011

BAZARIAN, J. **O problema da verdade**. São Paulo: Círculo do Livro, 1980. 251 p. .

BEATTY, J.; CHEN, A. **Visual Models for Software Requirements**. 1^a. ed. Redmond: Microsoft Press, 2012. 442 p. .

BECKER, K. Individual and Organisational Unlearning: Directions for futuro research. **International Journal of Organisational Behaviour**, v. 9, n. 7, p. 659–670, 2005. Disponível em: <www.usq.edu.au/extrfiles/business/journals/HRMjournal/IJOBHome.htm>. Acesso em: 1 mar. 2011

BELLOTTTO, H. L. **Arquivos Permantentes: tratamento documental**. 4^a. ed. Rio de Janeiro: FGV Editora, 2006. 320 p. .

BERMAN, F. Got Data? A Guide to Data Preservation in the Information Age. **Communications of the ACM**, v. 51, n. 12, 2008.

BERNARDES, I. P. **Como avaliar documentos de arquivos**. São Paulo: Arquivo do Estado, 1998. 89 p. .

BERNSTEIN, J. H. The Data-Information-Knowledge-Wisdom Hierarchy and its Antithesis. Proceedings North American Symposium on Knowledge Organization Vol 2. **Anais...2009**. Disponível em: <<http://dlist.sir.arizona.edu/2633/>>

BLACKLER, F. Knowledge, Knowledge Work and Organizations: An Overview and Interpretation. **Organization Studies**, v. 16, n. 6, p. 1021–1046, jan. 1995. Disponível em:

<<http://oss.sagepub.com/cgi/doi/10.1177/017084069501600605>>. Acesso em: 28 jan. 2011

BLUMENTRITT, R.; JOHNSTON, R. Towards a Strategy for Knowledge Management. **Technology Analysis & Strategic Management**, v. 11, n. 3, p. 287–300, set. 1999. Disponível em:

<<http://www.informaworld.com/openurl?genre=article&doi=10.1080/095373299107366&magic=crossref||D404A21C5BB053405B1A640AFFD44AE3>>. Acesso em: 9 mar. 2011

BROOKS, P. C. The selection of records for preservation. **The American Archivist**, v. III, n. 4, p. 221–234, 1940. Disponível em:

<<http://archivists.metapress.com/content/u77415458gu22n65/?p=076a6575ce6441b18a1e864abf1a3be1&pi=1>>. Acesso em: 20 nov. 2010

CAÑAS, A. J. et al. Concept Maps: Integrating Knowledge and Information Visualization. In: TERGAN, S.; KELLER, T. (Eds.). **Knowledge and Information Visualization**. Berlin: Springer Berlin Heidelberg, 2005. 205–219 p. .

CAPLAN, P. What Is Digital Preservation? **Library Technology Reports**, v. 44, n. 2, p. 7–9, 2008. Disponível em:

<<http://proxy.mul.missouri.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=lxh&AN=29439252&site=ehost-live&scope=site>>

CARTER, P. **Big Data Analytics: Future Architectures , Skills and Roadmaps for the CIO**. Singapore: IDC, 2011. 16 p. Disponível em:

<<http://www.sas.com/resources/asset/BigDataAnalytics-FutureArchitectures-Skills-RoadmapsfortheCIO.pdf>>. Acesso em: 19 out. 2012

CEGE - COMITÊ EXECUTIVO DE GOVERNO. **Padrão de Metadados do Governo Eletrônico**. Brasília : CEGE, 2010. Disponível em:

<<http://www.governoeletronico.gov.br/acoes-e-projetos/e-ping-padroes-de-interoperabilidade/padrao-de-metadados-do-governo-eletronico-e-pmg>>. Acesso em: 2 maio. 2013

CHELL, R. **Records Lifecycle v. Continuum**. Newcastle : Information and Records Management Society, 2007. Disponível em:
<<http://www.irms.org.uk/download/637>>

CHEN, P. P.-S. The Entity-Relationship Unified View of Data Model-Toward a Unified View of Data. **ACM Transaction on Database Systems**, v. 1, n. 1, p. 9–36, 1976.

CHEN, S.-S. The paradox of digital preservation. **Computer**, v. 34, n. 3, p. 24–28, mar. 2001. Disponível em:
<<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=910890>>

CHOO, C. W. The knowing organization: How organizations use information to construct meaning, create knowledge and make decisions. **International Journal of Information Management**, v. 16, n. 5, p. 329–340, out. 1996. Disponível em:
<<http://linkinghub.elsevier.com/retrieve/pii/0268401296000205>>

CLOONAN, M. V. The Preservation of Knowledge. **Library Trends**, v. 41, n. 4, p. 594–605, 1993.

COLLINS, H. M. The Structure of Knowledge. **Social Research**, v. 60, n. 1, 1993.

COLLINS, H. M. humans , machines , and the structure of knowledge. **Stanford Electronic Humanities Review**, v. 4, n. 2, 1995. Disponível em:
<<http://www.stanford.edu/group/SHR/4-2/text/collins.html>>. Acesso em: 9 mar. 2011

CONSTANTOPOULOS, P.; DRITSOU, V. An ontological model for digital preservation. DigCCurr2007 - \international Symposium in digital Curation. **Anais...**Chapel Hill: 2007. Disponível em:
<<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.129.1241>>. Acesso em: 25 out. 2010

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS - CCSDS. **Reference Model for an Open Archival Information System (OAIS), Recommended Practice, Issue 2**. Washington, D.C.: CCSDS/NASA, 2012. 135 p. Disponível em: <<http://www.ccsds.org>>. Acesso em: 19 nov. 2012

COOK, T. From the record to its context: The theory and practice of archival appraisal since Jenkinson. **S. A. Archives Journal**, v. 37, 1995.

COOK, T. What is Past is Prologue : A History of Archival Ideas Since 1898 , and the Future Paradigm Shift. **Archivaria**, n. 43, p. 17–63, 1997.
Disponível em:
<<http://journals.sfu.ca/archivar/index.php/archivaria/article/view/12175/13184>>

COOK, T. Archival science and postmodernism : new formulations for old concepts. **Archival Science**, v. 1, n. 1, p. 3–24, 2001.

COOK, T. **Appraisal Methodology: Macro-Appraisal and Functional Analysis**. Disponível em:
<<http://www.collectionscanada.gc.ca/government/disposition/007007-1035-e.html>>.

COOK, T. Macroappraisal in Theory and Practice: Origins, Characteristics, and Implementation in Canada, 1950–2000. **Archival Science**, v. 5, n. 2-4, p. 101–161, 12 set. 2006. Disponível em:
<<http://www.springerlink.com/index/10.1007/s10502-005-9010-2>>. Acesso em: 2 out. 2011

COUTURE, C. Archival Appraisal : A Status Report. **Archivaria**, v. 1, n. 59, 2005. Disponível em:
<<http://journals.sfu.ca/archivar/index.php/archivaria/article/view/12502/13624>>. Acesso em: 11 nov. 2010

DAVENPORT, T. H.; PRUSAK, L. **Working Knowledge: How Organizations Manage What They Know**. [s.l: s.n.]. 199 p. .

DAVID, P. A.; FORAY, D. Accessing and expanding the science and technology knowledge base. **STI Review**, n. 16, 1995.

DAVID, P. A.; FORAY, D. Una introducción a la economía y a la sociedad del saber. **Revista internacional de ciencias sociales**, n. 171, mar. 2002.

DAVIES, N. J. Knowledge management. **BT Technology Journal**, v. 18, n. 1, p. 62–63, 2000.

DIGITAL CURATION CENTRE. **What is digital curation?** Disponível em: <<http://www.dcc.ac.uk/digital-curation/what-digital-curation>>. Acesso em: 21 set. 2012.

DUARTE, J. Entrevista em profundidade. In: DUARTE, J.; BARROS, A. (Eds.). **Métodos e Técnicas de Pesquisa em Comunicação**. 2ª. ed. São Paulo: Atlas, 2006. 380 p. .

DURANTI, L. Rumo a uma teoria arquivística de preservação digital: as descobertas conceituais do Projeto InterPARES. **Arquivo & Administração**, v. 4, n. 1, p. 5–18, 2005.

DURANTI, L.; PRESTON, R. **International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2 : Experiential , Interactive and Dynamic Records**. Rome: Associazione Nazionale Archivistica Italiana, 2008. 828 p. .

DURANTI, L.; SUDERMAN, J.; TODD, M. **A Framework of Principles for the Development of Policies , Strategies and Standards for the Long-term Preservation of Digital Records**. Vancouver: The University of British Columbia, 2008. Disponível em: <[http://www.interpares.org/display_file.cfm?doc=ip2\(pub\)policy_framework_document.pdf](http://www.interpares.org/display_file.cfm?doc=ip2(pub)policy_framework_document.pdf)>. Acesso em: 16 nov. 2011.

EISENBERG, J. D. **OASIS OpenDocument Essentials Using OASIS OpenDocument XML**. Airlie Beach - Australia: Friends of OpenDocument Inc, 2005. 303 p. .

FAHEY, L.; PRUSAK, L. The Eleven Deadliest Sins of Knowledge Management. **California Management Review**, v. 40, n. 3, p. 265–276, 1998.

FERREIRA, M. **Introdução à Preservação Digital**: Conceitos, estratégias e actuais consensos. Guimarães: Escola de Engenharia da Universidade do Minho, 2006. 88 p. Disponível em: <<http://hdl.handle.net/1822/5820>>

FLECK, J. Contingent Knowledge and Technology Development. **Technology Analysis & Strategic Management**, v. 9, n. 4, 1997.

FRICKÉ, M. The knowledge pyramid : a critique of the DIKW hierarchy. **Journal of Information Science**, v. 35, 2009. Disponível em: <<http://jis.sagepub.com/content/35/2/131>>. Acesso em: 23 nov. 2010

GANTZ, J. et al. **IDC - The Expanding Digital Universe: A Forecast of Worldwide Information Growth Through 2010**. Framingham: IDC/EMC, 2007. Disponível em: <http://www.emc.com/about/destination/digital_universe/pdf/Expanding_Digital_Universe_IDC_WhitePaper_022507.pdf>. Acesso em: 16 out. 2010.

GANTZ, J.; REINSEL, D. **The Digital Universe Decade – Are You Ready ?** Framingham: IDC/EMC, 2010. Disponível em: <<http://idcdocserv.com/925>>. Acesso em: 9 nov. 2010.

GASSER, U.; PALFREY, J. **When and How ICT Interoperability Drives Innovation**. [s.l.] Berkman Center Research Publication, 2007. Disponível em: <<http://dx.doi.org/10.2139/ssrn.1033226>>

GLADNEY, H. M. **Preserving Digital Information**. 1. ed. Berlin: Springer-Verlag, 2007. v. 28314 p. .

GÓMEZ-PÉREZ, A. Ontological Engineering: A state of the art. **Expert Update: Knowledge Based Systems and Applied Artificial Intelligence**, v. 2, n. 3, 1999. Disponível em: <<http://oa.upm.es/6493/>>. Acesso em: 3 out. 2011

GOVERNMENT OF THE NORTHWEST TERRITORIES. **The Life Cycle of Records. Records Management Bulletin** Yellowknife : Department of Public Works and Services, 2002. Disponível em: <<http://www.pws.gov.nt.ca/records/bulletins.htm>>. Acesso em: 23 nov. 2010

GRADMANN, S. **INTEROPERABILITY**- A key concept for large scale , persistent digital libraries. : Briefing Paper[s.l.: s.n.]. Disponível em: <<http://www.digitalpreservationeurope.eu/publications/briefs/interoperability.pdf>>.

GRAHAM, P. S. **Intellectual Preservation** : Electronic Preservation of the Third Kind. , 1994. Disponível em: <<http://www.clir.org/pubs/reports/graham/intpres.html>>

GRUBER, T. Ontology. In: LIU, L.; ÖZSU, M. T. (Eds.). **Encyclopedia of Database Systems**. [s.l.] Springer-Verlag, 2009.

GRUBER, T. R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. **International Journal Human-Computer Studies**, v. 43, n. 5-6, p. 907–928, 1995. Disponível em: <<http://tomgruber.org/writing/onto-design.htm>>. Acesso em: 25 maio. 2011

GUPTA, B.; IYER, L. S.; ARONSON, J. E. Knowledge management: practices and challenges. **Industrial Management & Data Systems**, v. 100, n. 1, p. 17–21, 2000. Disponível em: <<http://www.emeraldinsight.com/10.1108/02635570010273018>>

HARVEY, R. **Preserving Digital Materials**. München, Germany: Saur Verlag, 2005. 246 p. .

HEDSTROM, M. Exploring the Concept of Temporal Interoperability as a Framework for Digital Preservation. Third DELOS Network of Excellence Workshop on Interoperability and Mediation in Heterogeneous Digital Libraries. **Anais...Darmstadt: ERCIM**, 2001. Disponível em: <<http://www.ercim.eu/publication/ws-proceedings/DelNoe03/10.pdf>>. Acesso em: 20 fev. 2013

HEDSTROM, M.; LEE, C. A. Significant properties of digital objects: definitions , applications , implications. DLM-Forum 2002. **Anais...Barcelona: INSAR - European Archives News**, 2002.

HIGGINS, S. The DCC Curation Lifecycle Model. **The International Journal of Digital Curation**, v. 3, n. 1, p. 134–140, 2008.

HOFMAN, H. Review: Some Comments on Preservation Metadata and the OAIS Model. **DigiCULT.Info - A Newsletter on Digital Culture**, n. 2, 2002. Disponível em: <<http://www.digicult.info/pages/index.php>>. Acesso em: 9 dez. 2012

HORRIDGE, M. et al. **A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools**. 1.1. ed. Manchester: The University Of Manchester, 2007. 103 p. .

IHMC. **Constructing Good Concept Maps**. Disponível em: <http://cmapskm.ihmc.us/servlet/SBReadResourceServlet?rid=1064009710027_279131382_27088&partName=htmltext>. Acesso em: 25 jan. 2013.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Functional Requirements for Bibliographic Records**. Haia: [s.n.]. Disponível em: <<http://www.ifla.org/VII/s13/frbr/>>.

INTERPARES PROJECT. **Creator Guidelines**: Making and maintaining digital materials. Vancouver: The University of British Columbia, 2007.

INTERPARES PROJECT. **Preserver Guidelines**: Preserving digital records - Guidelines for organizations. Vancouver: [s.n.].

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH (ICPSR). **Principles and Good Practice for Preserving Data**. : IHSN Working Paper[s.l.] IHSN, 2009. Disponível em: <<http://www.ihsn.org/home/download.php?file=IHSN-WP003.pdf>>.

JONES, M.; BEAGRIE, N. **Preservation Management of Digital Materials** : The Handbook. York, UK: Digital Preservation Coalition, 2008. 160 p. Disponível em: <http://www.dpconline.org/component/docman/doc_download/299-digital-preservation-handbook>. Acesso em: 15 maio. 2011

KAKABADSE, N. K.; KAKABADSE, A.; KOUZMIN, A. Reviewing the knowledge management literature: towards a taxonomy. **Journal of Knowledge Management**, v. 7, n. 4, p. 75–91, 2003. Disponível em: <<http://www.emeraldinsight.com/10.1108/13673270310492967>>. Acesso em: 2 jun. 2011

KASTELLEK, M. Practical Limits to the Scope of Digital Preservation. **Information Technology and Libraries**, v. 31, n. 2, 2012. Disponível em: <<http://ejournals.bc.edu/ojs/index.php/ital/article/view/2167/pdf>>

KENDAL, S.; CREEN, M. **An Introduction to Knowledge Engineering**. London: Springer-Verlag, 2007. 287 p. Disponível em: <<http://books.google.com/books?hl=en&lr=&id=9Z5XRoy3TQ8C&oi=fnd&pg=PA1&dq=An+Introduction+to+Knowledge+Engineering&ots=s5OUIT7Nr6&sig=Qi-UtSDX1maSb5xfDTeYfnlTbE4>>

KLEIN, P. D. Epistemology. In: ROUTLEDGE (Ed.). **Routledge Encyclopedia of Philosophy**. London: [s.n.].

KUNY, T. A Digital Dark Ages ? Challenges in the Preservation of Electronic Information. 63RD IFLA Council and General Conference. **Anais...**Copenhagen: IFLA, 1997. Disponível em: <<http://archive.ifla.org/IV/ifla63/63kunyl.pdf>>. Acesso em: 5 out. 2010

LAM, A. **Tacit Knowledge, Organisational Learning and Innovation: A Societal Perspective**. [s.l: s.n.]. Disponível em: <<http://www3.druid.dk/wp/19980022.pdf>>.

LAVILLE, C.; DIONNE, J. **A construção do saber**. Belo Horizonte: Editora UFMG, 1999. 340 p. .

LE BOEUF, P. Brave New FRBR World. 5th IFLA Meeting of Experts on an International Cataloguing Code. **Anais...**Pretória: IFLA, 2007. Disponível em: <[http://www.imeicc5.com/download/portuguese/Presentations2c_BraveNewFRBRWorld\(PR\)_Port.pdf](http://www.imeicc5.com/download/portuguese/Presentations2c_BraveNewFRBRWorld(PR)_Port.pdf)>. Acesso em: 24 nov. 2011

LEE, C. K.; FOO, S.; GOH, D. On the Concept and Types of Knowledge. **Journal of Information & Knowledge Management**, v. 5, n. 2, p. 151–163, 2006.

LEE, K. et al. The State of the Art and Practice in Digital Preservation. **Journal Of Research Of The National Institute Of Standards And Technology**, v. 107, n. 1, p. 93–106, 2002. Disponível em: <<http://introking3.com/002ftp/33/ase2/j71lee.pdf>>

LEVY, D. M. Fixed or Fluid? Document Stability and New Media. Proceedings of the 1994 ACM European conference on Hypermedia technology. **Anais...**ACM Press, 1994. Disponível em: <<http://portal.acm.org/citation.cfm?id=192760>>

LUDWIG, J. About the Complexity of a Digital Preservation Theory and Different Types of Complex Digital Objects (J.-P. Chanod et al., Eds.). Automation in Digital Preservation - Dagstuhl Seminar. **Anais...**Dagstuhl - Germany: Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany, 2010. Disponível em:

<<http://www.dagstuhl.de/Materials/Files/10/10291/10291.LudwigJens.ExtAbstract.pdf>>. Acesso em: 20 nov. 2012

LUNDVALL, B.-A. **The Social Dimension of The Learning Economy. Learning**[s.l: s.n.]. Disponível em: <<http://www3.druid.dk/wp/19960001.pdf>>.

LUNDVALL, B.-A. The economics of knowledge and learning. **Knowledge Creation Diffusion Utilization**, n. November, 2003. Disponível em: <http://www.emeraldinsight.com/Insight/ViewContentServlet?contentType=Book&Filename=/published/emeraldfulltextarticle/pdf/10_1016_s0737-1071_04_08002-3.pdf>

LYMAN, P.; VARIAN, H. **How Much Information ?** Berkeley: UC Berkeley, 2000. Disponível em: <<http://www2.sims.berkeley.edu/research/projects/how-much-info/how-much-info.pdf>>. Acesso em: 16 out. 2010.

LYMAN, P.; VARIAN, H. **How Much Information ? 2003**. Berkeley: UC Berkeley, 2003. Disponível em: <http://www2.sims.berkeley.edu/research/projects/how-much-info/2003/printable_report.pdf>. Acesso em: 16 out. 2010.

MACEVICIUTE, E. Long-Term Digital Preservation and Scholarly Communication. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 17, n. esp. 2, p. 1–18, 18 dez. 2012. Disponível em: <<http://www.periodicos.ufsc.br/index.php/eb/article/view/27162>>. Acesso em: 8 jan. 2013

MARTINS, D. C. M. Abordagem Sócio-Técnica Dora Cristina Moreira Martins. 2006. Disponível em: <<http://doramartins.com>>

MCKEMMISH, S. Yesterday , Today and Tomorrow : A Continuum of Responsibility. Records Management Association of Australia 14th National Convention. **Anais...Perth: RMAA**, 1997. Disponível em: <<http://www.infotech.monash.edu.au/research/groups/rcrg/publications/reco-rdscontinuum-smckp2.html>>

MCKEMMISH, S. Placing records continuum theory and practice. **Archival Science**, v. 1, n. 4, p. 333–359, dez. 2001. Disponível em: <<http://www.springerlink.com/index/10.1007/BF02438901>>

METS EDITORIAL BOARD. **Metadata Encoding & Transmission Standard**. Disponível em: <<http://www.loc.gov/standards/mets/>>. Acesso em: 1 set. 2013.

MILLAR, J.; DEMAID, A.; QUINTAS, P. Trans-organizational Innovation: A Framework for Research. **Technology Analysis & Strategic Management**, v. 9, n. 4, p. 1997, 1997.

MORAIS SQUIRRA, S. C. DE. A convergência tecnológica. **Revista Famecos**, p. 79–85, 2005. Disponível em: <<http://revistaseletronicas.pucrs.br/ojs/index.php/revistafamecos/article/viewFile/3324/2582>>. Acesso em: 16 ago. 2012

MOREIRA, M. A. **A teoria da aprendizagem significativa**. 1^a. ed. Brasília: Editora UnB, 2006. 186 p. .

MUGICA, M. M. M. Análisis de los modelos teóricos de gestión de los documentos eletrónicos: ciclo de vida y continuum. **Arquivo & Administração**, v. 4, n. 1, p. 31–39, 2005.

MUSGRAVE, A. **Common sense, science, and scepticism: a historical introduction to the theory of knowledge**. 1^a. ed. Cambridge: [s.n.]. 328 p. Disponível em: <<http://books.google.com.br/books?id=t-7Gvj4yHvYC&printsec=frontcover#v=onepage&q&f=false>>

NATIONAL LIBRARY OF AUSTRALIA. **Guidelines for the Preservation of Digital Heritage**. Organization Paris: Unesco, 2003. Disponível em: <<http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>>. Acesso em: 9 nov. 2010.

NISO - NATIONAL INFORMATION STANDARDS ORGANIZATION. **Understanding Metadata**. Bethesda: NISO Press, 2004. 20 p. Disponível em: <www.niso.org>. Acesso em: 25 maio. 2012

NONAKA, I.; TAKEUCHI, H. **Criação de Conhecimento na Empresa**. Tradução Ana Beatriz Rodrigues; Priscilla Martins Celeste. 2^a. ed. Rio de Janeiro: Campus, 1997. 358 p. .

NOVAK, J. D.; CAÑAS, A. J. The Origins of the Concept Mapping Tool and the Continuing Evolution of the Tool. **Information Visualization Journal**, v. 5, n. 3, 2006. Disponível em: <<http://ivi.sagepub.com/>>. Acesso em: 23 jun. 2008

NOVAK, J. D.; CAÑAS, A. J. **The Theory Underlying Concept Maps and How to Construct and Use Them**. Florida: Institute for Human and Machine Cognition, 2008. Disponível em: <<http://cmap.ihmc.us/Publications/>>. Acesso em: 20 fev. 2013.

NOY, N. F.; MCGUINNESS, D. L. **Ontology Development 101 : A Guide to Creating Your First Ontology**. Stanford: Stanford University, 2001. Disponível em: <http://protege.stanford.edu/publications/ontology_development/ontology101.pdf>.

O'DELL, C.; HUBERT, C. **The New Edge in Knowledge**: how knowledge management is changing the way we do business. Hoboken: Wiley, 2011. 236 p. .

OCLC/RLG WORKING GROUP ON PRESERVATION METADATA. **Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects**. Reproduction. Dublin: OCLC, 2002. Disponível em: <http://www.oclc.org/content/dam/research/activities/pmwg/pm_framework.pdf?urlm=161391>. Acesso em: 20 nov. 2012.

PINKETT, H. T. American Archival Theory : The State of the Art. **The American Archivist**, v. 44, n. 3, p. 217–222, 1981.

POPPER, K. R. **Conhecimento Objetivo**: Uma abordagem evolucionária. Tradução Milton Amado. 1a. ed. Belo Horizonte: Itatiaia, 1999. 394 p. .

PREMIS EDITORIAL COMMITTEE. **PREMIS Data Dictionary for Preservation Metadata - Version 2.2**. Washington, D.C.: Library of Congress, 2012. 272 p. Disponível em:

<<http://www.loc.gov/standards/premis/v2/premis-2-2.pdf>>. Acesso em: 20 nov. 2012

RHOADS, J. B. **The Role of archives and records management in national information systems** : a RAMP study. Paris: UNESCO, 1989. 72 p. Disponível em:

<<http://unesdoc.unesco.org/images/0008/000847/084735eo.pdf>>. Acesso em: 26 out. 2010

RIBEIRO, F. Os Arquivos na era pós-custodial: reflexões sobre a mudança que urge operar. IV Encontro de Outono - Memória, História e Patrimônio - Bibliotecas, Arquivos e Museus. **Anais...**Vila Nova de Famalicão: 2001.

RLG/OCLC. **Trusted Digital Repositories** : Attributes and Responsibilities. Mountain View, CA: Research Libraries Group, 2002. Disponível em:

<<http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf>>. Acesso em: 2 mar. 2013.

ROMHARDT, K. Processes of Knowledge Preservation : Away from a Technology Dominated Approach. 21st Annual Meeting of the German Society for Artificial Intelligence. **Anais...**Freiburg: DFKI - German Research Center for Artificial Intelligence, 1997. Disponível em:

<<http://www.dfki.uni-kl.de/~aabecker/Freiburg/Final/Romhardt/romhardt.html>>. Acesso em: 12 dez. 2009

ROUSSEAU, J.-Y.; COUTURE, C. **Os fundamentos da disciplina arquivística**. Tradução Magda Bigotte De Figueiredo. 1º. ed. Lisboa: Dom Quixote, 1998.

ROWLEY, J. The wisdom hierarchy : representations of the DIKW hierarchy. **Journal of Information Science**, v. 33, n. 2, p. 163–180, 2007. Disponível em: <<http://jis.sagepub.com/content/33/2/163>>. Acesso em: 23 jan. 2011

SAEKOW, A.; JIRACHIEFPATTANA, A. Towards a Pragmatic Methodology to Bridge the Gap in e-Government Interoperability : A Case of Patient Referral Information Exchange. Technical Symposium at ITU Telecom World (ITU WT). **Anais...**2011. Disponível em:

<<http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6100937&url=http>

://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6100937>. Acesso em: 1 mar. 2012

SAYÃO, L. F. Uma outra face dos Metadados: Informações para a gestão da preservação digital. **Bibli: Revista eletrônica de Biblioteconomia e Ciência da Informação**, v. 15, n. 30, p. 1–31, 2010.

SAYÃO, L. F. Repositórios digitais confiáveis para a preservação de periódicos eletrônicos científicos. **PontodeAcesso**, v. 4, n. 3, p. 68–94, 2010. Disponível em: <www.pontodeacesso.ici.ufba.br>

SAYÃO, L. F.; SALES, L. F. CURADORIA DIGITAL: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade: Estudos**, v. 22, n. 3, p. 179–191, 2012. Disponível em: <http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/12224/8586>. Acesso em: 13 ago. 2013

SCHELLENBERG, T. R. **Arquivos modernos**: princípios e técnicas. Tradução Nilza Teixeira Soares. 6^a. ed. Rio de Janeiro: FGV Editora, 2006. 386 p. .

SCHREIBER, G. et al. **Knowledge Engineering and Management**: The CommonKADS Methodology. Bradford: MIT, 2000. 408 p. .

SERRA, A. P. G. Convergência tecnológica em sistemas de informação. **Integração**, n. 47, p. 333–338, 2006. Disponível em: <ftp://ftp.usjt.br/pub/revint/333_47.pdf>. Acesso em: 16 ago. 2012

SERRA, J. S. Gestión de los documentos digitales: estrategias para su conservación. **El Profesional de la Información**, v. 10, n. 9, 2001.

SHARMA, N. **The Origin of the “Data Information Knowledge Wisdom” Hierarchy**. Disponível em: <http://www-personal.si.umich.edu/~nsharma/dikw_origin.htm>. Acesso em: 4 jul. 2008.

SILVA, A. M. DA; RIBEIRO, F. avaliação em arquivística - reformulação teórico-prática de uma operação metodológica. **Páginas a&b**, n. 5, p. 57–113, 2000.

SILVA, A. M. DA; RIBEIRO, F. A Avaliação de Informação : uma operação metodológica. **Páginas a&b**, n. 14, p. 7–37, 2004.

SILVA, A. M. DA; RIBEIRO, F. Perspectivar a avaliação como operação metodológica no âmbito da ciência da informação. IX Congress ISKO. **Anais...**Valencia: 2009.

SOUSA, R. T. B. DE. A classificação como função matricial do que-fazer arquivístico. In: **Arquivística: temas contemporâneos**. 2^a. ed. Distrito Federal: SENAC, 2008. 79–172 p. .

STUDER, R.; BENJAMINS, V. R.; FENSEL, D. Knowledge Engineering : Principles and methods. **Data & Knowledge Engineering**, v. 25, p. 161–197, 1998.

TAMBOURIS, E. et al. The role of interoperability in eGovernment applications: An investigation of obstacles and implementation decisions. 2008 Third International Conference on Digital Information Management. **Anais...**Ieee, nov. 2008. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4746798>>. Acesso em: 15 nov. 2012

THIBODEAU, K. Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. The State of Digital Preservation: An International Perspective. **Anais...**Washington: CLIR and Library of Congress, 2002. Disponível em: <<http://www.clir.org/pubs/reports/pub107/thibodeau.html>>. Acesso em: 11 ago. 2011

THOMASSEN, T. A First Introduction to Archival Science. **Archival Science**, v. 1, n. 4, p. 373–385, 2001.

TILLET, B. **O q u e F R B R ?** Washington : Library of Congress, 2003. Disponível em: <www.loc.gov/catdir/cpsio/o-que-e-frbr.pdf>. Acesso em: 25 maio. 2011

TILLET, B. **The FRBR Model**. Australian Committee on Cataloguing, 2004. Disponível em: <www.nla.gov.au/lis/stdnrd/grps/acoc/tillett2004.ppt>. Acesso em: 16 jun. 2011

TING, Z.; LI, D. Comparative Evaluation of Major IR Systems for Preservation. iPRES - 2007 International Conference on Preservation of Digital Objects. **Anais...**Beijing: 2007. Disponível em: <<http://rdd.sub.uni-goettingen.de/conferences/ipres07/>>

TOLK, A.; DIALLO, S. Y.; TURNITSA, C. D. Applying the Levels of Conceptual Interoperability Model in Support of Integrability , Interoperability , and Composability for System-of-Systems Engineering. **Journal of Systemics, Cybernetics and Informatics**, v. 5, n. 5, p. 65–74, 2007. Disponível em: <<http://www.iiisci.org/journal/sci/Contents.asp?var=&Previous=ISS7900>>. Acesso em: 18 ago. 2012

TOLK, A.; MUGUIRA, J. A. The Levels of Conceptual Interoperability Model. Simulation Interoperability Workshop. **Anais...**Orlando: Simulation Interoperability Standards Organization (SISO), 2003. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.196.1923&rep=rep1&type=pdf>>. Acesso em: 11 maio. 2012

UPWARD, F. Structuring the Records Continuum - Part One : Postcustodial principles and properties. **Archives and Manuscripts**, v. 2, n. 24, p. 268–285, 1996. Disponível em: <<http://www.infotech.monash.edu.au/research/groups/rcrg/publications/reco-rdscontinuum-fupp1.html>>

URDANETA, I. P. **Gestión de la Inteligencia, Aprendizaje Tecnológico Y Modernización Del Trabajo Informacional**. Caracas: Universidad Simon Bolivar, 1992. 253 p. .

WANG, W.; TOLK, A.; WANG, W. The Levels of Conceptual Interoperability Model : Applying Systems Engineering Principles to M & S. Spring Simulation Multiconference SpringSim09. **Anais...**San Diego: SCS, 2009. Disponível em: <<http://portal.acm.org/citation.cfm?id=1655398>>. Acesso em: 18 ago. 2011

WIIG, K. M. Knowledge management: Where did it come from and where will it go? **Expert Systems with Applications**, v. 13, n. 1, p. 1–14, jul. 1997. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0957417497000183>>

WILSON, A. **Significant Properties Report**. London, UK: King's College London, 2007. Disponível em:
<http://www.significantproperties.org.uk/wp22_significant_properties.pdf>. Acesso em: 12 nov. 2012.

XEXÉO, G. Big data: Computação para uma sociedade conectada e digitalizada. **Ciência Hoje**, v. 51, n. 306, p. 18–23, 2013.

YAMAOKA, E. J. Ontologia para mapeamento da dependência tecnológica de objetos digitais no contexto da curadoria e preservação digital. **AtoZ**, v. 1, n. 2, p. 65–78, 2012. Disponível em:
<<http://www.atoz.ufpr.br/index.php/atoz/article/view/23>>. Acesso em: 26 jul. 2013

YAMAOKA, E. J.; GAUTHIER, F. A. O. A Preservação Digital de Documentos Governamentais : Um Problema de Interoperabilidade ? 42 JAIIO - Jornadas Argentinas de Informática - 7º Simposio Argentino de Informatica En El Estado - SIE 2013. **Anais...**Córdoba - Argentina: SADIO, 2013.

YAMAOKA, E. J.; GAUTHIER, F. O. Ontologia de dependência tecnológica de documentos digitais: instrumento de apoio à preservação digital. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 17, n. esp. 2 - III SBCC, p. 211–226, 18 dez. 2012. Disponível em:
<<http://www.periodicos.ufsc.br/index.php/eb/article/view/27423>>. Acesso em: 25 set. 2013

YAMAOKA, E. J.; GAUTHIER, F. O. Objetos digitais: em busca da precisão conceitual. **Informação & Informação**, v. 18, n. 2, p. 77–97, 30 ago. 2013. Disponível em:
<<http://www.uel.br/revistas/uel/index.php/informacao/article/view/16162>>. Acesso em: 25 set. 2013

ZACK, M. H. Managing Codified Knowledge. **Sloan Management Review**, v. 40, n. 4, p. 45–58, 1999.

APÊNDICE A – QUESTÕES-CHAVE DAS ENTREVISTAS

A estrutura das entrevistas buscou investigar os seguintes aspectos do *framework*:

- Aplicabilidade do *framework*
- Eficácia do processo de seleção e as ações de preservação de objetos digitais
- Avaliação global do *framework* e alinhamento quanto ao objetivo proposto de preservação do conhecimento
- Identificação dos pontos fortes,
- Pontos de melhoria e sugestões para aprimoramento do *framework*

As questões utilizadas foram:

- 1) Considerando os componentes e os processos, o *framework* é aplicável nas organizações?
- 2) O processo definido de seleção dos objetos digitais atende o propósito de eliminar objetos desnecessários e aliviar os processos subsequentes?
- 3) Os componentes do grupo ações de preservação desempenham papéis relevantes no processo de superar a obsolescência tecnológica dos objetos digitais?
- 4) Quais os pontos fortes do *framework* proposto?
- 5) Que pontos melhorias/aprimoramento identificou?
- 6) Quanto ao atingimento do objetivo de preservar o conhecimento armazenado em objetos digitais, que grau de satisfação atribui? (zero a 10)

APÊNDICE B – PERFIL DOS ENTREVISTADOS

Entrevistado	Perfil	Formação e experiência
A	Tecnologia da Informação	<ul style="list-style-type: none"> • Engenheiro Industrial Eletricista - ênfase Eletrônica Industrial e Telecomunicação • Mestre em Engenharia Elétrica e Informática Industrial • Especialista em Formação Pedagógica do Docente de Nível Superior • Analista de Informática com especialização em desenvolvimento • Atuou em atividades de prospecção, projetos e suporte associados a TIC, em empresas como IBM, Banco HSBC e Siemens, em âmbito nacional e internacional. • Professor de IES – Graduação e Pós-graduação
B	Tecnologia da Informação	<ul style="list-style-type: none"> • Analista de Informática com especialização em desenvolvimento • Especialista em Tecnologia da Informação • Atua na coordenação e no desenvolvimento de <i>framework</i> para desenvolvimento de aplicações • Colaborador atuante em comunidades <i>opensource</i> • Contribuidor oficial do projeto <i>IceScrum</i>
C	Gestão do Conhecimento	<ul style="list-style-type: none"> • Doutor em Administração • Professor de IES – Graduação e Pós-graduação • Empresário • Consultor
D	Gestão do Conhecimento	<ul style="list-style-type: none"> • Engenheiro de Produção • Mestre em Ciência da Informação • Especialista em Gestão do Conhecimento • Analista de informática • Coordenou projetos de TI de abrangência nacional e internacional

		<ul style="list-style-type: none"> • Professor de IES – Pós-graduação
E	Preservação de documentos digitais	<ul style="list-style-type: none"> • Especialista em gestão e preservação de documentos arquivísticos digitais. • Mestre em informática • Pesquisador em projeto de pesquisa internacional em preservação digital • Coordena iniciativas e preservação documentos digitais de abrangência nacional
F	Preservação de documentos digitais	<ul style="list-style-type: none"> • Especialista em gestão, preservação de documentos de arquivos, patrimônio histórico e arquivos digitais. • Atua em iniciativas nacionais e internacionais relacionadas à arquivos e memória.

APÊNDICE C – SÍNTESE DAS RESPOSTAS

1. Considerando os componentes e os processos, o <i>framework</i> é aplicável nas organizações?	
Entrevistado	Resposta
A	Acho que sim. Pelo que vi está bastante abrangente e ao mesmo tempo está englobando aspectos que eu acho importante nessa questão de como dar conta da complexidade de alguns aspectos como a dinâmica do processo a medica que ocorre o envelhecimento do material, a questão da necessidade ou não da participação de pessoas no processo de decisão. Para mim o processo está claro e pertinente.
B	Acho que sim. Percebo que é o que tem faltado nas empresas que trabalhei. As soluções geralmente não tratam estes aspectos de avaliação e seleção e o resultado normalmente é 8 ou 80. Ou elimina tudo ou guarda tudo quando lixo e importante ficam juntos, depois ninguém consegue mais classificar "o que" é "o que". Geralmente a decisão é adquirir um sistema de GED ²¹ ou CMS ²² e tudo fica ali
C	Parece ser aplicável sim. Porém é necessário ser organizações de porte para organizar e operacionalizar esse <i>framework</i> . Considero que alguns tipos de organizações são mais propensas como bancos, organizações do sistema judiciário, pela questão legal, histórica e científica na área do Direito. É necessário porte para fazer o investimento e aplicação do <i>framework</i> .
D	Acho que sim. Por exemplo, a organização que trabalho tem a preocupação com documentos em papel, mas não temos nada no tratamento de documentos digitais. Portanto, a resposta é: sim, é aplicável nas organizações.
E	Acho que é aplicável, porém ele depende de um sistema bastante complexo, mas é o que defendemos no governo. O uso do documento na forma digital tem um custo. Como hoje não temos como fugir desse uso, precisamos arcar com esse custo. O <i>framework</i> é aplicável, mas terá um custo grande. Por outro lado, no caso do governo, poder preservar esses documentos pelo tempo necessário, para garantir os direitos do cidadão é preciso arcar com esse custo. A avaliação, pelo retorno que trará, acaba pagando esse custo. O difícil será

²¹ GED – Gestão Eletrônica de Documentos

²² CMS – Content Management System

	convencer o investimento, mas é uma necessidade inquestionável. Essa necessidade é mais clara nas grandes organizações. Numa pequena empresa em que o volume dos dados é pequeno, a gestão até é possível.
F	O documento digital não é uma entidade autônoma. Ele é criado, gerido, acessado por meio de um sistema. O documento digital precisa do sistema para existir. O <i>framework</i> abrange todo o conjunto de operações necessárias e ajuda também a reconfigurar os sistemas existentes. Um sistema de preservação é uma necessidade, considerando inclusive as imposições legais de mantê-los acessíveis e autênticos.

2. O processo definido de seleção dos objetos digitais atende o propósito de eliminar objetos desnecessários e aliviar os processos subsequentes?

Entrevistado	Resposta
A	Acho que sim. Principalmente no pronto de vista prático. A ideia de preservar tudo não é prática e talvez purista. Além disso, o exercício e o processo de valoração dos documentos se transformam num novo conhecimento, caso contrário, daqui a mil anos será necessário pegar uma massa imensa de informação para fazer essa valoração do zero. O próprio processo de seleção poderá formar uma nova base de conhecimento. O próprio critério de seleção dos colaboradores poderá ser insumo para mapear os valores de uma determinada época. É claro que na seleção ocorrerão perdas, mas é necessário buscar o equilíbrio com as questões práticas, ganhar de um lado e perder do outro.
B	Sim, pois geralmente o que ocorre é que o gestor decide o que é importante. Por exemplo, em uma empresa de desenvolvimento de <i>software</i> , guarda a documentação dos sistemas e joga o resto fora.
C	É uma questão que me incomoda. Quando falamos de dados como registros bancários, não vejo problema, mas em alguns aspectos de documentos de empresas e de governo, decidir o que é arquivável ou não, acho que esbarra também na questão ética. Quem é que vai julgar aquilo que vai ser guardado ou eliminado. A questão da frequência por exemplo. Posso ficar vinte anos sem usar ou acessar um objeto e depois desse período o acesso a esse objeto se tornar muito importante. Acho que

	<p>esse é o grande dilema, embora o processo de quarentena do <i>framework</i> possa ser uma válvula de escape para essa questão. Acho difícil a gente conseguir dimensionar um prazo, para essas situações, pois essa necessidade será sempre uma exceção. Se um documento ficou esquecido durante quarenta anos e alguém resolveu resgatar isso, então deve ser por um motivo muito importante. Não vejo muito motivo para resgatar um documento de 30 ou 40 anos a não ser que isso seja vital para a organização.</p> <p>Para o setor governo, preocupa-me que o grupo de colaboração para a avaliação não acabe virando um grupo sensor.</p> <p>Quando pensamos em preservação não podemos pensar em 100%, mas ao mesmo tempo quando falamos em descarte, quais são os critérios para fazer isso. Como evitar que dados politicamente sensíveis sejam eliminados?</p> <p>Mas acho que são questões críticas quando falamos de preservação, não do seu <i>framework</i>.</p>
D	<p>Acho que sim. O processo de avaliação e reavaliação e estratégia de manter os objetos em quarentena garante a eliminação dos objetos desnecessários. Porém, a preocupação é em relação ao que hoje é avaliado como desnecessário, mas que amanhã haveria uma reavaliação. O problema é o falso negativo, isto é, avaliar hoje algo como desnecessário, e no futuro descobrir que era necessário.</p>
E	<p>O processo de seleção proposto faz a avaliação dos documentos que não são arquivísticos e reavaliação dos documentos arquivísticos.</p> <p>No tocante aos documentos que não são arquivísticos é um ganho maravilhoso, pois isso não é realizado nas organizações. Os critérios utilizados para avaliação são os utilizados para elaboração da tabela de temporalidade.</p> <p>No caso dos documentos arquivísticos, o processo acaba dando uma chance de sobrevivência para aqueles que foram destinados à eliminação, o que é interessante, pois há a insegurança dos avaliadores quanto à decisão de eliminação.</p> <p>Com isso o processo permite a seleção mais controlada, mais consequente e permite a eliminação necessária.</p> <p>A eliminação com segurança é fundamental e o processo dá essa segurança.</p> <p>O que temos observado é a eliminação ocorrer por uma decisão de um gestor ou pelo acaso. Com o simples armazenamento dos dados em arquivos históricos, um dia perde-se o acesso. Portanto, a eliminação pode ocorrer por uma decisão arbitrária ou pelo acaso.</p>

	Esse processo de seleção permite a seleção consequente e isso é um ganho enorme.
F	Fica evidenciado o que é arquivístico e não-arquivístico e melhora a desempenho da avaliação.

3. Os componentes do grupo ações de preservação desempenham papéis relevantes no processo de superar a obsolescência tecnológica dos objetos digitais?

Entrevistado	Resposta
A	Acho importante, sem o gerenciamento da evolução tecnológica, o "gap" tecnológica pode se tornar algo terrível para lidar, mesmo que você guarde o objeto de alguma forma.
B	Não tenho muito conhecimento a respeito, mas acho que a informática é cíclica. Antes, o modelo era centralizado, ficou tudo distribuído e agora parece estar centralizando novamente.
C	Fiquei contente em conhecer o modelo de objetos digitais decomposta em físico, lógico e conceitual. O módulo de dependência tecnológica parece endereçar várias questões no tocante preservação do objeto.
D	Acredito que sim. De tudo que vi até agora, as estratégias de preservação são as quatro: preservação de tecnologia, emulação, encapsulamento e migração. Considerando que as ações de preservação consideraram essas estratégias, não conheço nada que pudesse ser acrescida.
E	É fundamental! Os sistemas hoje não preveem essas ações.
F	Sim. O acesso é o elemento chave. Sem as ações de preservação, temos os bits e com o tempo perdermos o acesso ao documento. O objeto estar inacessível é o mesmo que não tê-lo. Temos um artefato ininteligível. Um avião no museu não é um avião, é um artefato que parece um avião, pois o avião é aquele que voa. Então, se não temos mais acesso ao objeto, ele não é mais um documento. Não posso mais interpretar, utilizar e reutilizar.

4. Quais os pontos fortes do *framework* proposto?

Entrevistado	Resposta
A	A abrangência. Aparentemente ele supre todo o ciclo do processo de preservação de informações e conhecimento. A identificação de todas as fases e atores e a separação em

	módulos que facilita a evolução do processo, pode evoluir ou mudar alguns critérios ou processo, sem necessitar reformular todo o <i>framework</i> . A visão de buscar a partir do conceitual, algo aplicável.
B	Os processos que estão bem descritos. Aparecem aspectos ausentes nos atuais sistemas, o grupo de colaboração, a ontologia de dependência tecnológica.
C	A visão de superar a obsolescência da tecnologia parece ser um dos pontos fortes do <i>framework</i> e o processo contínuo de monitoramento e reavaliação é outro ponto forte da sua proposta.
D	É ele ser um <i>framework</i> voltado para documento digital. O ponto forte é ele existir. Outro ponto que considero forte é a estratégia da quarentena. Achei interessante preservar alguma coisa em que pese, por exemplo, o aspecto da frequência. Isto é, que o simples fato do uso frequente, mesmo que em outros critérios não seja qualificado, leva a olhar o objeto com mais atenção. A existência de um repositório da quarentena torna-se uma salvaguarda para os casos de falso negativo, especialmente nos casos de documentos históricos.
E	Acho que são os dois processos das perguntas anteriores: o processo de avaliação e seleção e as ações de preservação. Embora a Arquivologia venha trabalhando com a avaliação há tempos, esse processo não tem sido incorporado nos sistemas. Quando você traz o processo de avaliação e seleção, inclusive para documentos não-arquivístico, acho isso um ganho enorme. Se os sistemas de informações conseguirem incorporar esse processo de seleção mais controlado e consequente, isso será fundamental, pois as organizações poderão preservar o que é importante, ou pelo menos aquilo que é mais importante. Sempre ocorrerão perdas, mas a seleção é fundamental. E a ação de preservação, pois nada adiantará a seleção se não conseguir manter aquilo que foi selecionado. Os pontos fortes são esses dois processos.
F	Sim, se não houver a seleção e as ações de preservação, você não terá as "duas pernas" da preservação. O processo de reavaliação permitirá conhecer cada vez melhor o que você tem na organização e dá mais qualidade naquilo que está avaliado.

5. Que pontos melhorias/aprimoramento identificou?

Entrevistado	Resposta
A	<p>Acho que a migração de tecnologia um grande desafio. Os processos de migração não são infalíveis. Como garantir que um objeto foi migrado corretamente? Será necessário um operador humano para verificar isso? Pode requerer no futuro, até uma disciplina que estude e garanta as transformações confiáveis, como testar.</p> <p>No processo de avaliação e seleção, soluções para objetivar os fatores subjetivos. A própria capacitação de uma comunidade e não influenciar o resultado da seleção é um aspecto complexo.</p>
B	<p>Senti ausência do processo do gestor. Quem é o responsável pelo documento num determinado estado? O <i>framework</i> poderia explicitar a necessidade de existir uma área de Gestão do Conhecimento que ficaria responsável pelo <i>framework</i>. Com a minha visão mais técnica, de implementação, alguns processos a exemplo da quarentena, poderia ser mais detalhado.</p>
C	<p>Pelo que vi a proposta cobre os pontos críticos pra solucionar o problema. Parece-me que a proposta está bem sustentada para esse propósito de acessibilidade de longo prazo.</p> <p>O problema da seleção, eu não considero uma sugestão de melhoria, pois é um problema humano. O problema de "quem" vai decidir o que é lixo ou não, não é do <i>framework</i>, é um dilema ético. Preocupa-me a possibilidade da influência política e ética da seleção. Eu acho que o problema está além da dimensão do <i>framework</i>. Se fosse possível, um aprimoramento seria tornar o <i>framework</i> imune a essas questões.</p>
D	<p>Acho que a questão da assinatura digital. Abordar como os documentos protegidos por certificados digital serão tratados e como dar a garantia de autenticidade do documento digital sem usar o certificado digital. Acho que é um ponto que deve ser abordado, inclusive sendo excluído do <i>framework</i> se for o caso.</p> <p>A segunda, seria a questão do falso negativo. Verificar se de fato o esquema da quarentena daria a salvaguarda para esses casos.</p>
E	<p>A questão da cópia. Necessita ver em que contexto está. Se a cópia for de instituições diferentes, isto é, pertencente a fundos diferentes, neste caso são documentos diferentes.</p> <p>Por exemplo, um ofício que é emitido pelo órgão A para o órgão B, embora sejam objetos digitais iguais, no fundo do</p>

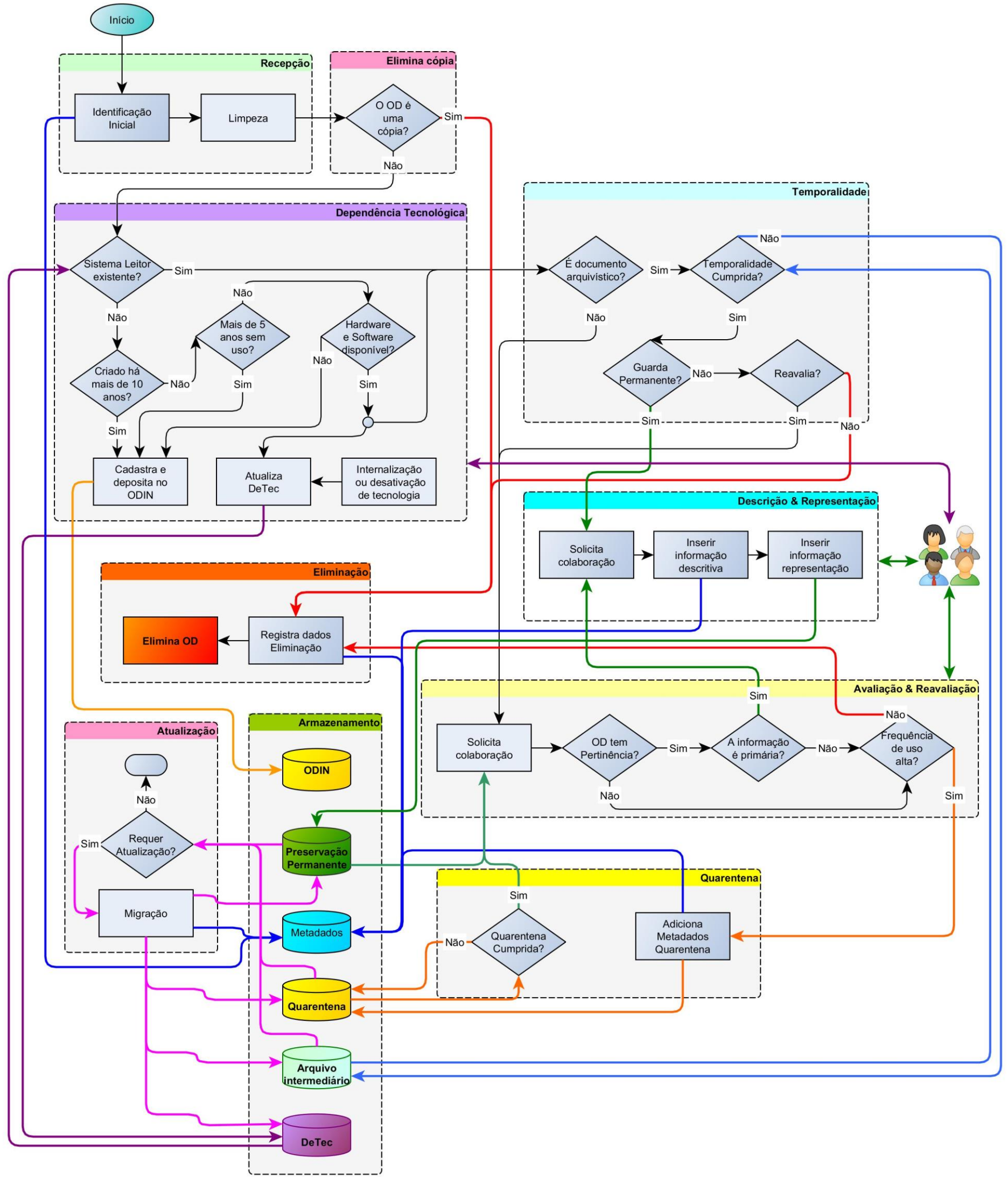
	<p>órgão A será arquivado como documento enviado e o órgão B, em seu fundo arquivará esse ofício como correspondência recebida.</p> <p>O documento é único em cada atividade que participa. Às vezes, o mesmo documento integra dossiês diferentes. Em cada um dos dossiês ele é único.</p> <p>Portanto, o checksum teria que ser conjugado com os metadados para verificar se são documentos diferentes.</p>
F	<p>A questão da identificação da cópia. O contexto e o status do documento podem gerar objetos tecnologicamente idênticos, mas que são dois documentos. As cópias podem pertencer a documentos diferentes. As repetições desnecessárias são as duplicatas.</p> <p>Dependendo da arquitetura do sistema de informação essas cópias podem não ocorrer. O mesmo documento pode apenas conter as relações, isto é, as informações como quem emitiu e quem foi o destinatário.</p>

6. Quanto ao atingimento do objetivo de preservar o conhecimento armazenado em objetos digitais, que grau de satisfação atribui? (zero a 10)

Entrevistado	Resposta
A	Acho que vai ao encontro da gestão/construção de conhecimento. No trabalho atual e passado percebi que sofremos muito pela falta de preservação de conhecimento. Você percebe retrabalho o tempo todo. Acho excelente a proposta. Daria 10.
B	Só não daria 10, pelo meu desconhecimento profundo do assunto, mas daria 9.
C	Eu diria que está nos pontos superiores. O <i>framework</i> prevê todos os aspectos, preocupa-se com questões importantes. De modo geral o <i>framework</i> está bem completo. Certamente acima de 8.
D	O <i>framework</i> garante o acesso tratando a questão da obsolescência tecnológica, o cuidado de estar alinhada com o critério de temporalidade da arquivística, a quarentena para salvaguardar os objetos para futuras mudanças de avaliação. Sim, o <i>framework</i> atinge o objetivo de preservar os conhecimentos armazenados nos objetos. Eu atribuiria 9 por não ter tratado nada sobre certificados digitais.
E	Do ponto de vista conceitual está atingindo. Contempla tudo que temos discutido. Numa primeira visão, como uma proposta conceitual, daria 10. Gostei muito, achei

	interessante.
F	10. Acho que tudo que é necessário está contemplado. A Identificação do objeto, a seleção, a preservação, os diferentes tipos de armazenamento. As chances dos objetos serem preservados estão contempladas dentro dos princípios que temos defendido. As incertezas também estão contempladas.

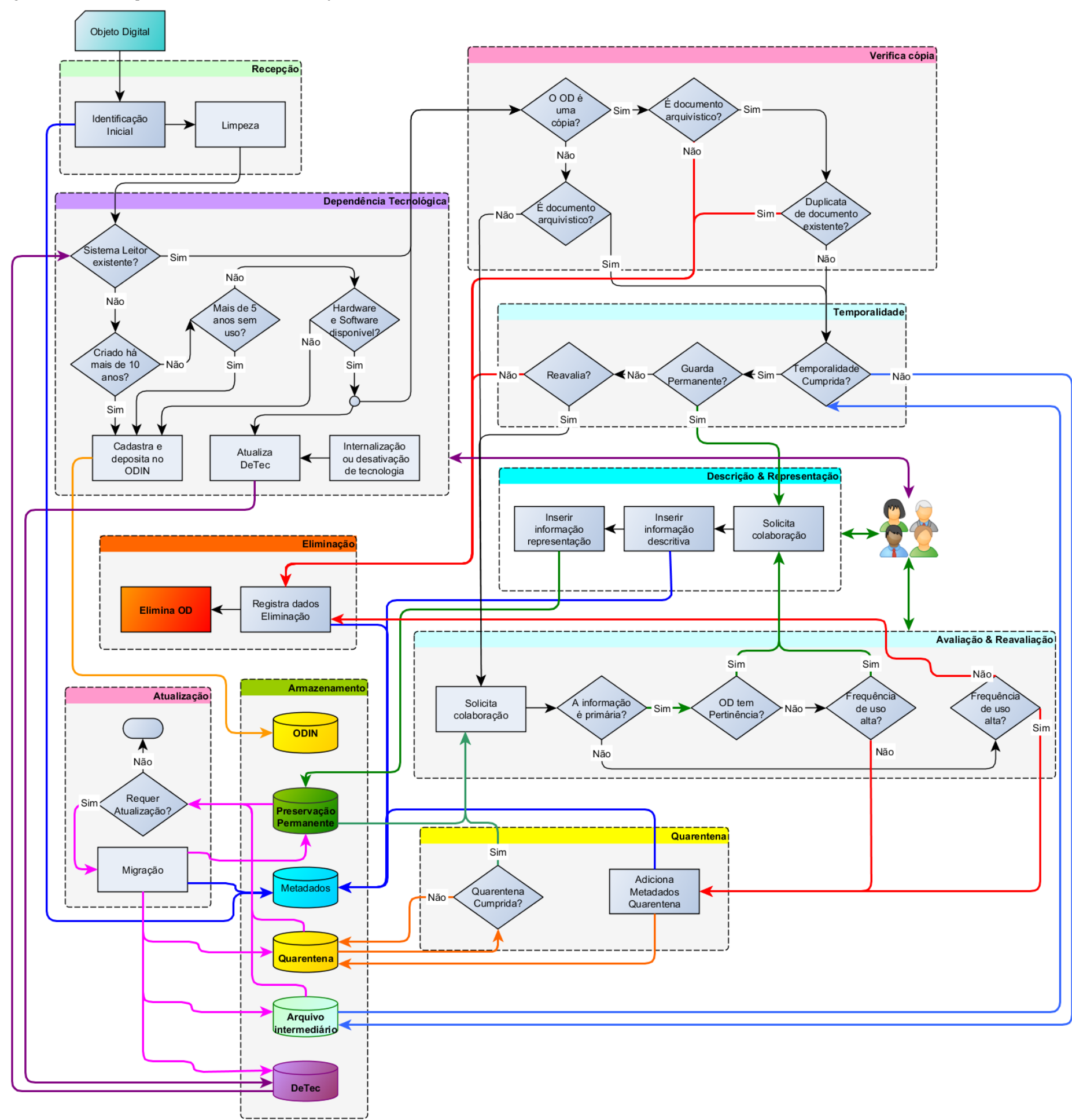
APÊNDICE D – VISÃO GERAL DO FRAMEWORK
Figura 57 – Visão completa detalhada do Framework



Fonte: Elaborado pelo autor

APÊNDICE E – VISÃO GERAL DO FRAMEWORK AJUSTADO

Figura 58 – Visão completa detalhada do *Framework* Ajustado



Fonte: Elaborado pelo autor

